Naturewatch Canada: Metadata analysis for a citizen-science based monitoring program

By
François L'Ecuyer

Thesis submitted to the School of Graduate Studies and Research in partial fulfilment of the requirements for the Degree of Master of Science, Geography

Department of Geography, Environment and Geomatics
University of Ottawa
February 23rd 2017

Abstract

NatureWatch Canada, a citizen science program, collects and analyses data pertaining to plant phenology, frog species and ice coverage over water bodies in Canada to monitor trends through time and space in relation to climate change. An important question is whether this database is currently usable to infer environmental changes through space and time. This thesis presents a metadata analysis of the Nature Watch database in order to identify the spatial validity, quality, reliability and usability of the current data. We first explore citizen science through a review of the literature, followed by a detailed analysis of the content of the database. We also produce an example Newsletter for each module to illustrate some of the current trends in the data. The Frogwatch Newsletter shows how weather conditions in 2001 may have favored a population spurt of Leopard frogs resulting from more spawning ponds essential for this species. Next the Plantwatch Newsletter reveals that an increase in Aspen poplar and Prairie crocus in 2002 may have been due to fires during the preceding years. Lastly, the Icewatch Newsletter shows how ice formation, but not melt, reveals a clear trend of occurring 17 days later over the last 100 years with a stronger increase starting in the sixties for Western Ontario. In summary, this thesis presents a detailed metadata analysis of the Naturewatch database in order to provide recommendations for its improvement in the future. Improving programs like Naturewatch Canada is important to monitoring climate and ecological changes that could be applied throughout the Canadian North which are not currently well represented in this database.

Acknowledgements

I wish to thank my thesis supervisor Dr. André Viau, without whose continued guidance I would have been unable to complete this project. I also wish to thank Dr. Michael Sawada for teaching me how to use GIS programs which have served me well in producing this thesis.

I would like to thank Frank Oliva for his help during the early stages of the project, I could not have accessed the data I needed, or figured out how to work with it, without your assistance.

I would also like to think Chantal Arcand and Nathalie Maras, for helping me get organized and directing me to do the paperwork for my Teaching Assistant positions which allowed me to fund my Masters'.

I am very thankful for the support of my parents, Dr. Michel L'Ecuyer and Hélène Gagnon, in pursuit of my studies and this project.

I would also wish to extend thanks to my committee members Dr. Robert McLeman and Dr. Jackie Dawson, for their suggestions and feedback which allowed me to complete this thesis in a timely fashion.

I would like to also thank all of my friends for their support throughout my thesis studies by helping me to recharge my batteries when I needed a slight break.

Table of contents

Abstract ii

Acknowledgements iii

List of figures vii

List of tables xii

Chapter 1. Introduction 1

1.1 Objectives 1

1.2 Literature review 3

1.3 History of citizen science programs in North America 5

1.4 Characteristics of citizen science programs 6

1.5 History and traits of NatureWatch 10

1.6 Summary of literature review 14

Chapter 2. Methodology 16

2.1 Objective 1: Metadata analysis 16

2.1.1 Table/variable description 16

2.1.2 Data acquisition 19

2.1.3 Basic statistics 21

2.2 Objective 2: Data analysis 22

2.2.1 Spatial analysis 22

2.3 Objective 3: Example newsletters 23

Chapter 3. Frogwatch analysis 24

3.1 Frogwatch newsletter 33

Chapter 4. Plantwatch analysis 40

4.1 Plantwatch newsletter 48

Chapter 5. Icewatch analysis 59

5.1 Icewatch newsletter 65

Chapter 6. Summary 73

6.1 General conclusions 73

6.2 Frogwatch 74

6.3 Plantwatch 75

6.4 Icewatch 76

6.5 Recommendations for future work 77

References 80

Appendix A: Frogwatch analysis 84
Appendix A-1 Frogwatch tables 84
1.1 Frogwatch_observation 84
1.2 location 85
1.3 Observer 86
1.4 Frog_species 88
1.5 frogwatch_observation_code 88
1.6 Partner 88
1.7 age_range 89
1.8 habitat_type 89
Appendix A-2 Frogwatch trend analysis 89
Appendix A-3: Frogwatch spatial analysis 116
Appendix B: Plantwatch analysis 131
Appendix B-1: Plantwatch tables 131
1.1 plantwatch_observation 131
1.2 Location 131
1.3 Observer 132
1.4 Visit 134
1.5 plant_species 135
1.6 plantwatch_observation_code 135
1.7 partner 135
1.8 age_range 136
1.9 watch_programme 136
Appendix B -2: Plantwatch trend analysis 137
Appendix B-3: Plantwatch spatial analysis 169
Appendix B-4: Plantwatch observer analysis 189
Appendix C: Icewatch analysis 191
1.1 icewatch_observation 191
1.2 location 191
1.3 observer 192
1.4 icewatch_observation_code 194
1.5 Partner 194
1.6 age_range 195
Appendix D: Wormwatch tables 196
1.1 wormwatch_observation 196

1.2 location 197
1.3 observer 198
1.4 visit 199
1.5 Soil_type 200
1.6 Soil_colour 200
1.7 Worm_species 200
1.8 partner 200
1.9 age_range 201
Appendix E: Other Naturewatch tables 202
1.1 Administrator 202
1.2 admin_login_attempts 202
1.3 Content 203
1.4 qc_codes 203
1.5 taxonomy 203
1.6 validate_email 204
1.7 valid_parameter_code 204

List of figures

Figure 1: NatureWatch entity relationship diagram. 17
Figure 2: Example format of table description for the Frogwatch_observation table, and its variables. 18
Figure 3: Example histogram of Frogwatch entry count by year. 21
Figure 4: Example mapping of the point density of Spring Peeper entries in Ontario. 22
Figure 5: Frogwatch yearly entries from 1996 to 2015. 25
Figure 6: Frogwatch observation locations. 26
Figure 7: Percentage of entries each year for the American Toad and the total entries. 27
Figure 8: Map of the data point density for American Toad entries. 28
Figure 9: Yearly percentage of the total number of entries for the aggregate and those for the Boreal chorus frog. 29
Figure 10: Boreal chorus frog entry density. 30
Figure 11: Yearly percentage of Bullfrog entries and Aggregate. 31
Figure 12: Map representing the entry density for bullfrog data points. 32
Figure 13: Leopard frog. 33
Figure 14: Spring peeper. 34
Figure 15: Yearly percentage for the Spring peeper and aggregate datasets. 36
Figure 16: Yearly percentage of total entries for the Leopard frog and aggregate from 1996 to 2015. 37
Figure 17: Map of all locations of Plantwatch events. 41
Figure 18: Total number of Plantwatch entries for each year from 1996 to 2015. 42
Figure 19: Yearly percentage of the total number of entries for the aggregate and those for Aspen poplar. 43
Figure 20: Map of the area density of Aspen Poplar entries. 44
Figure 21: Yearly percentage of the total number of entries for the aggregate and those for Saskatoon. 45
Figure 22: Saskatoon entry density map. 46
Figure 23: Yearly percentage of the total number of entries for the aggregate and those for Dandelion. 47
Figure 24: Dandelion entry density over Canada. 48
Figure 25: Dandelion flower in bloom. 50
Figure 26: Dandelion yearly percentage of total entries compared to the aggregate from 1996 to 2015. 51
Figure 27: Common purple lilac blooming flowers. 52
Figure 28: Common purple lilac and aggregate yearly percentage of total entries, from 1996 to 2015. 52
Figure 29: Wild strawberries bearing fruit. 53
Figure 30: Yearly percentage of the total entries for Wild strawberries and the aggregate from 1996 to 2015. 54
Figure 31: Aspen polar flowers in the early bloom stage. 55
Figure 32: Yearly entry percentage for Aspen poplar and the aggregate from 1996 to 2015. 56
Figure 33: Prairie crocus yearly entry percentage, compared with the aggregate, from 1996 to 2015. 57
Figure 34: Map of Icewatch observation locations throughout Canada. 61
Figure 35: Ice-off average ordinal dates for every year of Icewatch. 63
Figure 36: Ice-on average ordinal dates for every year of Icewatch. 64
Figure 37: Icewatch module of Naturewatch Canada. 65
Figure 38: Ice-on average ordinal dates for available years for Ontario in Icewatch. 66
Figure 39: Ontario ice-on average yearly ordinal date, only years with continuous data. 67
Figure 40: Ontario ice-off yearly ordinal date. 68
Figure 41: Map of the locations from the selected regions. 69
Figure 42: Yearly mean ordinal date for time of ice melt for the selected region. 70
Figure 43: Ice-on yearly average ordinal date for the selected region. 71

Figure 44: This graph shows the percentage of the total number of entries for each year for the Frogwatch_observation table. 91
Figure 45: Percentage of entries each year for the American Toad and the aggregate. 92
Figure 46: Yearly percentage of the total number of entries for the aggregate and those for Blanchard's cricket frog. 93
Figure 47: Yearly percentage of the total number of entries for the aggregate and those for the boreal chorus frog. 94
Figure 48: Yearly percentage of the total number of entries for the aggregate and those for the boreal toad. 95
Figure 49: Yearly percentage for the Bullfrog and aggregate. 96
Figure 50: Canadian toad and aggregate yearly percentage. 97
Figure 51: Yearly percentage for the Columbia spotted frog and aggregate. 98
Figure 52: Cope's treefrog and aggregate yearly aggregate. 99
Figure 53: Yearly percentages for Fowler's toad and aggregate. 100
Figure 54: Yearly percentages for the Gray treefrog and aggregate. 101
Figure 55: Great Basin spadefoot and aggregate yearly percentages. 102
Figure 56: Yearly percentages for the Great Plains toad and aggregate. 103
Figure 57: Green frog and aggregate yearly percentages. 104
Figure 58: Yearly percentages for the Mink frog and aggregate. 105
Figure 59: Leopard frog and aggregate yearly percentages. 106
Figure 60: Pacific treefrog and aggregate yearly percentages. 107
Figure 61: Yearly percentages for the aggregate and Pickerel frogs. 108
Figure 62: Plains spadefoot and aggregate yearly percentages. 109
Figure 63: Red-legged frog and aggregate yearly percentages. 110
Figure 64: Yearly percentages for Oregon spotted frog and aggregate. 111
Figure 65: Percentage of total entries by year, for the Spring peeper entries and the aggregate. 112
Figure 66: Western chorus frog and aggregate yearly percentages. 113
Figure 67: Yearly percentages for the aggregate and Tailed frog. 114
Figure 68: Wood frog and aggregate yearly percentages. 115
Figure 69: American toad entry density map. 117
Figure 70: Blanchard's cricket frog entry density map. 117
Figure 71: Boreal chorus frog entry density map. 118
Figure 72: Entry density map for the Boreal toad. 118
Figure 73: Bullfrog entry density map. 119
Figure 74: Entry density map for Canadian toad. 119
Figure 75: Columbia spotted frog entry density map. 120
Figure 76: Cope's treefrog entry density map. 120
Figure 77: Entry density map for Fowler's toad. 121
Figure 78: Gray treefrog entry density map. 121
Figure 79: Great Basin spadefoot entry density map. 122
Figure 80: Great plains toad entry density map. 122
Figure 81: Entry density map for the Green frog. 123
Figure 82: Leopard frog entry density map. 123
Figure 83: Mink frog entry density map. 124
Figure 84: Entry density map for the Pacific treefrog. 124
Figure 85: Pickerel frog entry density map. 125
Figure 86: Entry density map for the Plains spadefoot. 125
Figure 87: Red-legged frog entry density map. 126

Figure 88: Spring peeper entry density map. 126
Figure 89: Entry density map for the Tailed frog. 127
Figure 90: Western chorus frog entry density map. 127
Figure 91: Wood frog entry density map. 128
Figure 92: Graph of the percentage of total observations made for each count contribution. 129
Figure 93: Frogwatch aggregate entry counts by year. 137
Figure 94: Yearly percentage of the total number of entries for the aggregate and those for Aspen poplar. 137
Figure 95: Yearly percentage of the total number of entries for the aggregate and those for Bearberry. 138
Figure 96: Yearly percentage of the total number of entries for the aggregate and those for Northern bedstraw. 139
Figure 97: Yearly percentage of the total number of entries for the aggregate and those for Blue-head lily. 140
Figure 98: Yearly percentage of the total number of entries for the aggregate and those for bluets. 141
Figure 99: Yearly percentage of the total number of entries for the aggregate and those for Bunchberry. 142
Figure 100: Yearly percentage of the total number of entries for the aggregate and those for Sagebrush buttercup. 143
Figure 101: Yearly percentage of the total number of entries for the aggregate and those for Choke cherry. 143
Figure 102: Yearly percentage of the total number of entries for the aggregate and those for Cloudberry. 144
Figure 103: Yearly percentage of the total number of entries for the aggregate and those for Coltsfoot. 145
Figure 104: Yearly percentage of the total number of entries for the aggregate and those for Cranberry. 146
Figure 105: Yearly percentage of the total number of entries for the aggregate and those for Dandelion. 147
Figure 106: Yearly percentage of the total number of entries for the aggregate and those for Labrador tea. 148
Figure 107: Yearly percentage of the total number of entries for the aggregate and those for Larch. 149
Figure 108: Yearly percentage of the total number of entries for the aggregate and those for Common purple lilac. 150
Figure 109: Yearly percentage of the total number of entries for the aggregate and those for Lodgepole pine. 151
Figure 110: Yearly percentage of the total number of entries for the aggregate and those for Arctic lupine. 151
Figure 111: Yearly percentage of the total number of entries for the aggregate and those for Red Maple. 152
Figure 112: Yearly percentage of the total number of entries for the aggregate and those for Mayflower. 153
Figure 113: Yearly percentage of the total number of entries for the aggregate and those for Mountain avens. 154
Figure 114: Yearly percentage of the total number of entries for the aggregate and those for Prairie crocus. 155
Figure 115: Yearly percentage of the total number of entries for the aggregate and those for Queen’s cup. 156

Figure 116: Yearly percentage of the total number of entries for the aggregate and those for Rhodora. 156
Figure 117: Yearly percentage of the total number of entries for the aggregate and those for Saskatoon. 157
Figure 118: Yearly percentage of the total number of entries for the aggregate and those for Prickly saxifrage. 158
Figure 119: Yearly percentage of the total number of entries for the aggregate and those for Purple saxifrage. 158
Figure 120: Yearly percentage of the total number of entries for the aggregate and those for Star-flowered Solomon's seal. 159
Figure 121: Yearly percentage of the total number of entries for the aggregate and those for Starflower. 160
Figure 122: Yearly percentage of the total number of entries for the aggregate and those for Wild Strawberry. 161
Figure 123: Yearly percentage of the total number of entries for the aggregate and those for Sweetgale. 162
Figure 124: Yearly percentage of the total number of entries for the aggregate and those for Trillium. 162
Figure 125: Yearly percentage of the total number of entries for the aggregate and those for Twinflower. 163
Figure 126: Yearly percentage of the total number of entries for the aggregate and those for White water lily. 164
Figure 127: Yearly percentage of the total number of entries for the aggregate and those for Wolf willow. 164
Figure 128: Yearly percentage of the total number of entries for the aggregate and those for Early blue violet. 165
Figure 129: Yearly percentage of the total number of entries for the aggregate and those for Golden bean. 166
Figure 130: Yearly percentage of the total number of entries for the aggregate and those for Yarrow. 166
Figure 131: Yearly percentage of the total number of entries for the aggregate and those for Weeping forsythia. 167
Figure 132: Yearly percentage of the total number of entries for the aggregate and those for Paper birch. 168
Figure 133: Map of the entry density for Aspen poplar. 169
Figure 134: Entry density for Paper birch for Plantwatch. 169
Figure 135: Weeping forsythia entry density map. 170
Figure 136: Entry density map for Yarrow. 170
Figure 137: Golden bean entry density map. 171
Figure 138: Early violet entry density map. 171
Figure 139: Entry density map for Wolf willow. 172
Figure 140: Water lily entry density map. 172
Figure 141: Twinflower entry density map. 173
Figure 142: Entry density map for Trillium. 173
Figure 143: Map of entry density for Sweetgale. 174
Figure 144: Strawberry entry density map. 174
Figure 145: Starflower entry density map. 175
Figure 146: Entry density map for Solomon's Seal 175
Figure 147: Purple saxifrage entry density map. 176
Figure 148: Prickly saxifrage entry density map. 176

Figure 149: Entry density map for Saskatoon. 177
Figure 150: Rhodora entry density map. 177
Figure 151: Queen's cup entry density map. 178
Figure 152: Entry density map for Prairie Crocus. 178
Figure 153: Mountain avens entry density map. 179
Figure 154: Mayflower entry density map. 179
Figure 155: Red maple entry density map. 180
Figure 156: Arctic lupine entry density map. 180
Figure 157: Lodgepole pine entry density map. 181
Figure 158: Entry density map of Purple lilacs. 181
Figure 159: Larch entry density map. 182
Figure 160: Labrador tea entry density map. 182
Figure 161: Entry density map for the Dandelion. 183
Figure 162: Cranberry entry density map. 183
Figure 163: Coltsfoot entry density map. 184
Figure 164: Entry density map for Cloudberry. 184
Figure 165: Choke cherry entry density map. 185
Figure 166: Sagebrush buttercup entry density map. 185
Figure 167: Entry density map for Bunchberry. 186
Figure 168: Bluets entry density map. 186
Figure 169: Bluehead lily entry density map. 187
Figure 170: Entry density map for Northern bedstraw. 187
Figure 171: Bearberry entry density map. 188
Figure 172: Aspen poplar entry density map. 188
Figure 173: Percent of total entries originating from users with a given entry contribution. 189

List of tables

Table 1: Ten oldest entries for the Ontario FrogWatch, starting with March 15th, 1999. 19

Table 2: Example query in the Search tab of the NatureWatch database for the frogwatch_observation table. In this case we are searching for the entries where the species was identified as species 21. 20

Chapter 1. Introduction

Citizen science is a type of research with a fairly long history but which is only recently beginning to be used more extensively in large-scale academic research. The rise in popularity of this method of data collecting can be attributed to the acknowledgment of the limitations of conventional research methodologies when it comes to very large sample sizes and spatial-temporal distributions. This is particularly true for ecological and biological studies such as those for plant phenology, along with those of interest to the humanities (who better to report on superficial social and cultural data than the subjects themselves?). Historically citizen science programs tended to be started and run entirely by private citizens or groups investigating a matter of personal concern or interest. A shift in this trend is also being observed with larger governing bodies, such as the government of Canada in our case, turning to this method both as a means of connecting with the people under their jurisdiction and democratising science.

NatureWatch, a citizen science project currently under the stewardship of the University of Ottawa and Wilfrid Laurier University is a project aimed at acquiring data over wide areas on plant species composition through PlantWatch, ice cover of lakes via IceWatch, and amphibian populations with FrogWatch. There is also a WormWatch module, aimed for the most part at raising awareness in and interest for this project among youth. The following presents a metadata analysis of the NatureWatch databases. The goal of this study is to assess the current state of observations available in different modules (eg Frogwatch) in order to evaluate the quality, quantity, reliability, spatial-temporal distribution and trends in the data. This analysis will aid the NatureWatch program coordinators in decision-making and future directions.

1.1 Objectives

Objective 1: Conduct a metadata analysis of the NatureWatch database. This is done in order to better identify the possible issues with the data itself (uncertainties, errors related to data type and encoding per se). Naturally this concerns the sources, dates of entry, and possible interpretations of some of the questions contributors had to answer to fill out the database.

Objective 2: Analyse each of the NatureWatch modules (Frogwatch, Plantwatch, Icewatch) in turn. The analysis goes in depth with a description of spatial-temporal distribution and trends in the data.

Objective 3: Produce a series of three example newsletters (Frogwatch, Plantwatch, Icewatch) based on results from individual modules from the data analysis objective (2).

The thesis is presented as a traditional chapter-based thesis with the following chapters: Chapter 1 presents a thorough literature review on citizen science, and the research objectives; chapter 2 presents the general methodological approach; chapters 3, 4, 5 presents the results for the frogwatch, plantwatch and icewatch modules including a newsletter and chapter 6 provides a summary, conclusions and recommendations.

1.2 Literature review

Before describing the NatureWatch program, one of the questions we need to answer is whether citizen science can be a useful tool for research. In short, what is citizen science and why use the data derived from it?

We must first define what citizen science is. What the analysis leading to this review has made clear is that the understanding of the concept, while widespread, has many variations, and referring to it as "an increasingly ubiquitous term used to refer to a time-honored, evolving practice" (Crain *et al.*, 2014) can sum up the large number of recent research on the subject and perspectives regarding to it. Citizen science is the research method, or phenomenon, of allowing citizens to participate in a research project in various roles. The following definition aptly sums up the possible interpretations and variations of citizen science:

"Citizen science is typically understood as scientific activities—usually the collection and sometimes the processing of data—carried out by nonprofessional scientists in the context of a scientific project" (Stevens *et al.*, 2014)

This definition excludes private enthusiast initiatives which, as noted by Devictor *et al.* (2010), could be considered to be the original form of science (and citizen science) in the field of biogeography, because these where not organised by or coordinated with a research effort specifically designed to make the data accessible neither to the scientific community nor general public. There are several terms which, while not actually synonymous, are frequently used interchangeably with citizen science, such as crowd-sourcing and community-based monitoring. This is a consequence of the fact that citizen science programs have existed for much longer than the methodological studies on them (Tulloch *et al.*, 2013), meaning that the vocabulary is not yet fully established and often defined differently for each study (Zhao and Zhu, 2014).

There are several reasons for using citizen science data. First it can help fill in the gaps of conventional research with regards to data monitoring on large scales (spatial and temporal). Another important reason is that citizen science can be seen as an effort to democratize science – and in fact it is promoted as such (Mansell, 2014).Professional or academic research, while important to decision making on governmental scales, can seem distant and sometimes even disconnected to the wider public, which leads some to feel alienated or unconcerned by its findings. This could, in many ways, reflect a relatively recent shift in attitudes towards decision makers typified by a feeling of lack of true transparency or dialogue between decision makers and the people they nominally represent, at both the private and public level. Involvement of citizens may reconnect research and people in a way that could be highly beneficial to the immediate research goal and the broader acceptance and reliance on solid research by the people (Mansell, 2014). It must be noted however that it is not entirely certain that this would be the result of citizen science programs, as there are currently no universal means of measuring "citizen empowerment" (Mansell, 2014).

This last point is particularly relevant because it promotes involvement through the learning process and participation (Hiller, 1991). These are good things for any research but a complex issue like climate change benefits the most from a shift in participant attitude. The nature of climate change, being its causes (including ubiquitous human activity, i.e. done by everyone), its temporal scale (previous century with ongoing consequences of a magnitude variable based on current changes) and difficulty in mitigation for the individual tend to cause resignation and apathy towards it, as it is already underway and its inertia is/seems unstoppable. This attitude must be countered for any meaningful change to be effected. Therefore research that promotes involvement has the power to affect decisions and has a central place in the overall mitigation and adaptation strategies to climate change.

1.3 History of citizen science programs in North America

The last few decades have seen a boom in the number of citizen science programs thanks to information technologies, with the Internet taking pride of place in making the programs accessible and contribution from multiple disparate sources possible at all times (Tulloch *et al.*, 2013). This shift may also be the result of a change in perspective, or paradigm (Tippins and Jensen, 2012). As Tippins and Jensen (2012) mention, the gradual acknowledgement of the complexity and "unknowability" (the authors use the term chaos) of the world, particularly when we consider ecosystems and interconnected climatic cycles, leads one to seek complementary avenues for finding answers. In this case, one of these "new" avenues is citizen science. "Though sometimes thought of as a new idea, some CBM organizations have been monitoring ecosystems (and ecosystem components) for decades (i.e., Christmas Bird Count since 1900(Audubon 2008) and the British Trust for Ornithology (BTO) for over 50 years)." (Conrad and Hilchey, 2011)

The huge number of new citizen science programs in recent years has led some to consider this to be a recent field, however as the above quote points out, this is not the case. The principles of citizen science can be found in a great variety of earlier projects and research, even though these did not have the complete characteristics to make them actual citizen science projects. Indeed some fields in science, notably biogeography and phenology, were originally started by enthusiasts and amateur scientists (Devictor *et al.*, 2010), and were thus closer to citizen science than they were to traditional research.

For the purposes of this literature review, we have determined that the first citizen science initiative in North America is the Christmas bird count program started by the National

Audubon Society (Tulloch *et al.*, 2013 – Tippins and Jensen, 2012) in 1900. This program is the oldest project mentioned in the literature (in several articles as referenced) which undoubtedly conforms to the definition of a citizen science program. It is also ongoing, incredibly, and can also serve in an analysis of citizen science programs over time. Moreover, its longevity makes this data highly valuable in an historical context on bird species.

Another reason for considering the Audubon program the first true citizen science program is noted in Devictor *et al.* (2010). Unlike many other private initiatives of the past, the data for the Audubon Bird Count, even the earliest, is still available. Here we come to a crucial element of the definition or criteria we use for citizen science: not only must the data collection involve citizens, but the data itself must also be accessible to the wider public.

1.4 Characteristics of citizen science programs

As the previous section illustrated, citizen science as a research method is well established, but the scientific community has only begun to analyse its characteristics recently. An important question to ask then is what leads researchers to turn to crowd sourcing or citizen science data. The following section will highlight particular theoretical and anecdotal traits of citizen science inferred from its historical use and conceptual analyses from the literature.

The most easily observed common feature of citizen science programs is that they use "very simple standardized protocols replicated across many surveyed plots to draw broad conclusions across large spatial extents and/or temporal scales" (Devictor *et al.*, 2010). This illustrates one of the main advantages of data drawn from such sources: broad geographical scale through the high number of sample sources and the possibility to monitor for long periods of time, a common limitation of traditional research. Therefore, the citizen science approach can specifically address such a concern by employing a large number of contributors for a minimal investment (Tulloch *et al.*, 2013). As such, we should expect citizen science programs to crop up when a team wishes to analyse large-scale or even global phenomena: the best example of this, currently, is that of global warming impacts, which requires observations of "hot spots" such as glaciers, but also widespread data over every continent and locale to be inclusive.

When we state that the contribution of non-professional researchers is valuable for large-scale datasets, the temporal aspect is critical. For example, one article on astronomy made an important point: "Where amateurs really become indispensable is when time-series observations must be made" (Henden, 2011). The point here is that citizen-collected data can continue to monitor a given phenomenon longer than would be feasible for a research team, where time constraints can severely limit the observations.

While it's usually considered that citizen science allows for long-term monitoring, it is also useful for quickly gathering a large amount of data on sudden events (such as natural disasters) (Foody, 2014). This is because it can take advantage of the flush of interest and commitment associated with people's reaction to a particular disaster. Given the short time frame, however, is when the data's quality is most likely to be called into question because there is little time to establish a good methodology and ensure the earnestness of the contributions.

While citizen science has a keen edge over traditional research when it comes to carrying out large samplings at reasonable cost, it does have some shortcomings inherent to its nature. The first and most obvious is that employing non-professionals introduces a distinct bias in the data, depending on what is being monitored. In species composition research, citizen scientists will naturally tend to report highly visible species, meaning those which are overt and/or easily identified, over those species which are timid, stealthy or not clearly recognizable (Devictor *et al.*, 2010). The data will thus over-represent these "charismatic" species and under-represent less well-known species. This is also observable in the field of ornithology, a field where citizen science is well established (for instance with the Audubon Christmas bird count), where the most conspicuous species are the most looked for and observed (Tulloch *et al.*, 2013). Even if these concerns are addressed, and research indicates that there are clear ways to do so, there remains an issue of trust when it comes to the data's validity (Thornton and Leahy, 2012).

With the question of bias comes a concern about plain malicious contribution, in other words purposeful errors made by individuals or small groups with an interest in falsifying or otherwise discrediting the research (Foody, 2014), even if that interest is simple mischief ("joking"). This is particularly simple to conceptualise in the present, where numerous online campaigns become plagued by spam and agents known commonly as "trolls", but besides looking for clearly aberrant data points and reacting appropriately there are few ways to prevent this. The following quote sums up the problem: "Critically, in many cases, the end-user of the derived data has little or no control over the contributors and has no or negligible information on data quality" (Foody, 2014). Conceptually, it is reasonable to assume a project which appears "in the spotlight" of multiple media is most vulnerable to this phenomenon, particularly when the subject polarises the public.

In other words:

"Although crowdsourcing works on the principle that 'two heads are better than one', sometimes a crowd can return a vast amount of noise that may be of little relevance (Keen 2007)"

This is inherent to having a multitude of data contributors over a wide area, since verification cannot be done in any affordable way besides establishing a "hierarchy" of contributors among larger and better known actors, such as universities in the case of

NatureWatch. This may solve some issues but also runs against the democratising spirit of citizen science.

Another question is how to evaluate the individual contributions because individual citizen scientists are not equal and one may be more susceptible to bias or incomplete entries than another (Crain *et al.*, 2014). This can lead to serious issues since any method to address bias invariably depends on what bias exists in the contributors, which varies based on who they are and their qualifications (Edwards, 2014).

A critical long-term issue is that of participant retention. Essentially, the onset of a new citizen science program can usually expect some degree of enthusiasm from its non-professional participants. However, the initial interest in the program will almost invariably die out unless incentive and promotion plans are carried out (Zhao and Zhu, 2014).

The main objectives of volunteer monitoring data are management, awareness, education, serendipity, recreation, recreation, social and economic research, ecological knowledge and improving methods (Tulloch *et al.*, 2013). Seeing how these connect to the motivations of the participants is crucial to knowing how to encourage further participation and how to appeal to a larger audience.

Categorising contributors may also help to determine the data's reliability. Developing general incentives strategies may prove less effective than those aimed at the main contributors (Zhao and Zhu, 2014). Zhao and Zhu (2014) list 3 types of participants: super-contributors, contributors and outliers. In this research, the authors conclude that identifying what strategy is known to be successful in crowdsourcing initiatives may be the key to finding what outreach can most attract and retain involvement. (Stewart *et al.*, 2010, Zhao and Zhu, 2014).

Very few sources addressed the interdisciplinary nature of citizen science, instead focussing on its use for a particular discipline (Crain *et al.*, 2014). The same article goes on to suggest that employing citizen scientists should be taken as an opportunity not only to collect ecological data but also social and cultural data on the participant which could be combined to add a human layer to the data and potentially make correlations between culture and ecological perspective (Crain *et al.*, 2014). In short, there is a point to be made for collecting data both from (ecological data) and about (social/cultural data) of the participants (Crain *et al.*, 2014). This also flows well with the paradigm change from a mechanistic view of natural systems to a very complex interconnected one as mentioned earlier (Tippins and Jensen, 2012).

1.5 History and traits of NatureWatch

NatureWatch officially began in 2000 as a collaboration between Environment Canada and Nature Canada. Nonetheless, the literature hints at earlier origins from provincial programs, at least as far as the different modules (Plantwatch, Icewatch and Frogwatch) are concerned.

The first “contributor” program mentioned in the literature is the Alberta Plantwatch phenology program, as noted in this quote “The first large-scale Canadian phenology observer network started in Alberta in 1973. This decade-long survey of bloom dates of wild plants was initiated through the Federation of Alberta Naturalists” (Beaubien and Hamann, 2011). This program is linked to the eventual development of Alberta Plantwatch, then Prairie Plantwatch before being expanded in 2002 into the Canada Plantwatch.

Based on the literature (Beaubien and Hamann, 2011), and also on the descriptions found in the NatureWatch website (revamped in November 2014, to consolidate its change of direction from Environment Canada to the University of Ottawa in 2011), the Plantwatch segment of NatureWatch is primarily a flowering phenology program, and thus most data entries would be made in spring, but also throughout the other seasons except winter. What is immediately obvious is that the flowering time, combined with the species, is one of the most important variables to consider for detecting climate trends over time (Beaubien and Hamann, 2011), which is why winter is excluded as no plant flowers in winter conditions in Canada. Something to consider, however, is that for those plants still present and visible in winter (mainly large shrubs and trees), the composition could also serve as indicators over the longer term of local environmental conditions (which may reflect global changes). These changes would typically operate over much longer periods given both the lifespan and time needed to grow these plants, so the data collected in winter on these species will only become relevant after several more years or decades of data has been collected.

In its early days, the program was promoted by the distribution of booklets in 1988 to "potential observers", which included Universities (Beaubien and Hamann, 2011). The program would possibly have returned to its original size and purpose when Environment Canada cut the program's funding, but it could also have vanished altogether, which feeds into the recruitment/motivation aspect of citizen science.

Dr. Beaubien from the University of Alberta, an important contributor to the Alberta Plantwatch, makes notes on the participant retention of other programs compared to Plantwatch: "A study of motives for long-term participation by 150 volunteers in an 'Adopt-a-Stream' program revealed the following as most important: enjoying learning, helping the environment, feeling needed, having time for reflection, and benefitting from a well-organized program with good leadership (Ryan *et al.*, 2001). No specific studies of PlantWatch volunteers' motivations have yet been published." (Beaubien and Hamann, 2011). In short, for the most part, the analysis of participants has effectively not been conducted at any point in the program's history.

As mentioned above, citizen science programs rely on easily-reproducible methodologies to be used by the contributors (Devictor *et al.*, 2010), for example, accurate and consistent species identification is essential for Plantwatch and Frogwatch. We should thus expect there to be easily usable identification aids accessible to the public for those wishing to participate – and there is. The NatureWatch website does include pages for each monitored species.

Another important consideration is the governing structure of the program, which can be either consultative/functional, collaborative, or transformative (Conrad and Hilchey, 2011). Since the main standard of consultative governance is being government run (Conrad and Hilchey, 2011), we can conclude that NatureWatch does not fall in this category, as it is run by the University of Ottawa and assisted by several other ENGOs such as the David Suzuki Foundation, Nature Canada and Wilfrid Laurier University. At first glance, it would appear that NatureWatch is a collaborative program given its involvement of many stakeholder groups, being run from an intermediate level (as opposed to strictly top-down or grassroots). This would indicate the decision power of NatureWatch is promising, as this type of program was found to be relatively effective compared to the others cited (Conrad and Hilchey, 2011). It must be borne in mind that this classification changes based on when in its history the program is considered, as we might have concluded that it was a consultative/top-down program had we conducted the analysis a few years ago, when it was still under the purview of Environment Canada.

This shows the complexity of trying to categorise such a program over the years, as both its managers and contributors change, leading to changes in the methods and directions taken by the program. The one thing we can conclude with certainty is that NatureWatch was not a grassroots program, meaning it was not started by citizens. While the original programs were not necessarily developed in response to a particular perceived need, the “umbrella program” is itself mostly devoted to identifying and understanding the trends of responses to climate changes (Beaubien and Hamann, 2011). This is quite relevant as citizen science is generally well suited to issues of large scale, a description which suits climate change research in both time and space. What should the NatureWatch data hint at, in terms of trends? The literature, while not specific to this program, did present phenological studies made in Europe (Menzel *et al.*, 2005, 2006) which showed a linear trend of flowering dates/dates of first births or egg laying (for animals) and the shortening length of the winter season (Menzel *et al.*, 2006). This is a trend which would be expected to occur in this data as well if NatureWatch data is a good indicator of climate. If no such trend is observed, then the variables or perhaps the species employed are not good indicators.

Nonetheless, the question remains: is NatureWatch Canada capable of delivering these large-scale datasets? The following sums up an observation of plant phenology data for North America:

"However, there is a notable lack of phenology data for western Canada and Alaska, where the change in the spring warming signal over the last 50 years has been most pronounced globally (Rosenzweig *et al.*, 2007)."

The literature (Beaubien and Hamann, 2011, Rosenzweig *et al.*, 2007) indicates the data is mostly concentrated in Nova Scotia, the Prairies, Southeastern Ontario and Southwestern Quebec.

This illustrates another conceptual shortcoming of citizen science, one not mentioned in any detail in the literature: data derived from citizen science can only be expected to provide any volume of data in populous regions with contact between the general population and the research team or database. For remote regions, some contributions can be made but not necessarily reliably due to poor services in communication services for example. Areas with a small number of possible contributors comprise a large part of Canada (in fact the majority of its surface area), notably in higher latitudes which is where the impacts of climate change are more pronounced (IPCC, 2013). In short, while citizen science is a good approach for large-scale problems, it is only viable for large areas where people are distributed more or less evenly. This is an important issue for NatureWatch who must contend with promoting and outreaching in areas of low density such as those of the northern latitudes of the country.

Another consideration is that of incentive strategy, which is largely lacking for NatureWatch Canada. The “incentive” is intrinsic (Zhao and Zhu, 2014) because it relies on the desire to contribute to resolving a problem in a meaningful way or learn more on the issue of climate change. The best way to cultivate this intrinsic incentive is to encourage the data’s use in research by as many stakeholders as possible and to make available these articles to the NatureWatch website and reinforce its meaningfulness from the perspective of the participant. This avoids the problem encountered in research involving public (not limited to citizen science) data drawn from participants being “whisked away to pages of professional journals and government reports” (Chavis *et al.*, 1983). As we can see however, this incentive is rather indirect and its effectiveness might be limited.

1.6 Summary of literature review

Over the course of this literature review, we have determined that citizen science is the use of citizen-collected data for use towards a given research goal. We have also identified the advantages this field held for research on phenomena operating at a large scale, both temporally and spatially, such as climate change. These advantages come from the involvement of a large number of contributors spread over a large area and potentially continued monitoring over long periods of time, for example, the case for the Christmas Bird Count of the National Audubon Society. There are also limitations associated with involving the general public into research, as this can compromise the rigor of the methodology of data collection and casts doubt on the results' validity. This is an important consideration for NatureWatch Canada and on the general subject of climate change research where higher latitude areas (i.e. more sensitive to climate change impacts) may be under-represented in the database. This thesis explores the spatial and temporal distribution of the citizen data collected in order to propose recommendations going forward.

Chapter 2. Methodology

2.1 Objective 1: Metadata analysis

In the first objective, a detailed look at the available data from NatureWatch is documented. This includes data in all tables and for each module. This step is important in getting to know what data is used in objective 2 from data found insufficient to draw any significant conclusions. This work is done using ArcGIS, python and R software packages.

2.1.1 Table/variable description

There are several tables in the database, however only three are used for the data itself. The remaining tables include some to associate a number code (for species) to an English and French common name and the actual scientific identifier, as an example. The end product for this section of the analysis includes an entity relationship diagram, meaning a graphical representation of each table as a text box with a list of the columns and with the relationships symbolized as arrows and identified with codes to describe the nature of said relationship.

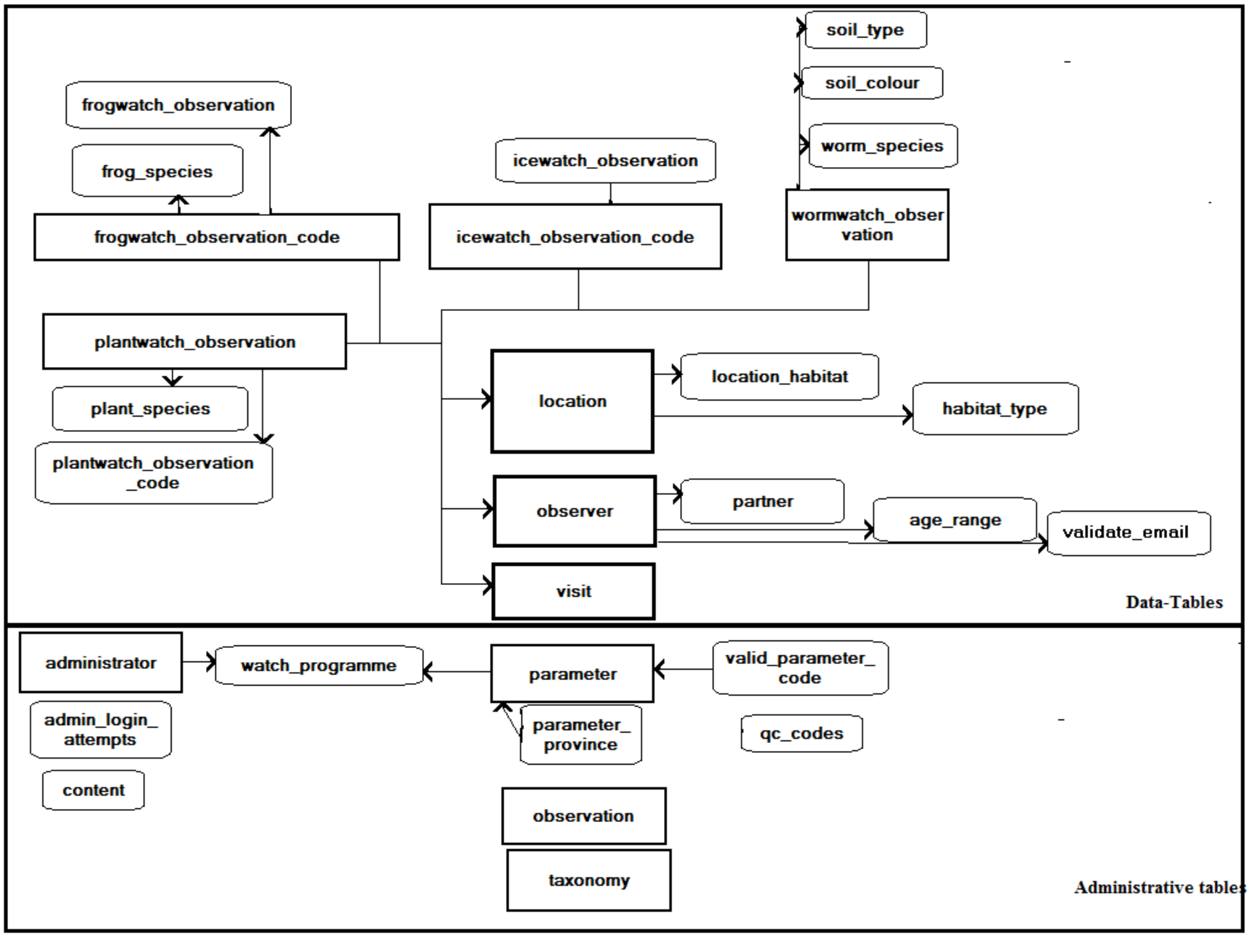

Figure 1: NatureWatch entity relationship diagram.

As for the format by which each table is described, each of these is numbered, and assigned a subsection in the appropriate module's appendix. Within this subsection we find a small sample of the table and its headings, along with a short paragraph describing both its purpose and which other table in the database references it (or which table it references). Figure 2 illustrates an example for the frogwatch_observation table. Essential references to the Wormwatch and Administrative tables are described in Appendix D and Appendix E, respectively.

()NatureWatch tables

().1 frogwatch_observation

One of the three primary data tables in NatureWatch. References the Frog_species, Location, frogwatch_observation_code and Location_habitat tables.

Table:frogwatch_observation			
Column heading	Data type	Reference type	Reference table
Pk frogwatch_observation	Integer		
Fk_frog_species	Varchar (2)	BTREE	Frog_species
. . .	. . .	. . .	. . .

Pk frogwatch observation: Arbitrary data entry identification number.

Fk_frogspecies: 2-digit integer identifying the species number. Links to the Frog_species table.

Figure 2: Example format of table description for the Frogwatch_observation table, and its variables.

The link to the frog_species table allows the database to assign a text equivalent (in French and English) to the 2-digit species code. Also note that the primary key (pk) for frogwatch observations connects to no other table. The pk identifier stands for primary key, meaning it is primary data (data which refers to itself only), and so every other variable preceded with pk shares this trait. Those preceded by fk (Foreign key) are by definition connected to another table (hence the term foreign).

For the analysis, the first thing we consider is the date of the first entries, which will determine to a certain extent what kind of predictions we can make from the data (the timespan of the dataset being very important for trend analysis). Using the Sort and Filter tool in Excel, we can sort the entries by date of observation.

FID_	Observatio	Frog_ID	Common_nam
1454	15/03/1999 0:00	1	American Toad
1456	15/03/1999 0:00	15	Leopard Frog
1470	15/03/1999 0:00	17	Pickerel Frog
1473	15/03/1999 0:00	3	Boreal Chorus Frog
1594	24/03/1999 0:00	24	Wood Frog
6	30/03/1999 0:00	21	Spring Peeper
1554	07/04/1999 0:00	21	Spring Peeper
1555	07/04/1999 0:00	15	Leopard Frog
1577	07/04/1999 0:00	24	Wood Frog
1457	08/04/1999 0:00	13	Green Frog

Table 1: Ten oldest entries for the Ontario FrogWatch, starting with March 15th, 1999.

As we can see, none of the observations for FrogWatch in Ontario date further than 16 years at the time of writing. This does entail that very long-term predictions and inferences cannot be made with this data alone as it does not go sufficiently far into the past. For mid or short term predictions it is enough, however, considering the short generation time for most of these species when compared to late-succession trees for instance (where many don't even begin reproducing before the second decade), which might pose a problem if the PlantWatch data does not go further back.

2.1.2 Data acquisition

The exact method for data acquisition depends on whether we are accessing data as a user on the NatureWatch website or as an administrator for the database itself. This is because some of the data are listed as, for example, an endangered species and cannot be accessed by the general public for reasons of conservation.

To extract specific data, this database uses the SQL language of the phpMyAdmin database program. To find information in the database we can use the headings for each table provided by the designer. For example, a "Search" functionality can facilitate certain queries by generating a SQL table (Table 2).

Column	Type	Collation	Operator	Value
fk_frog_species	Varchar(2)	Utf8_general_ci	=	21

Table 2: Example query in the Search tab of the NatureWatch database for the frogwatch_observation table. In this case we are searching for the entries where the species was identified as species 21.

The SQL code equivalent of this search is:

SELECT * FROM `frogwatch_observation` WHERE `fk_frog_species`=21. Using the SQL format is a fairly streamlined process, and also rather intuitive and self-explanatory for most simple operators. Both the inbuilt search function of the database and SQL allow for multiple-column searches with various operators. Effective use of this method requires a basic understanding of what certain symbols mean (notably the *, which means "everything" or "all"). The entry translates in common language as "display all of the rows in the frogwatch_observation table where the value in the fk_frog_species column is 21".

Obviously for research purposes this information is dubious because we do not know immediately what frog species is coded as 21 in the fk_frog_species column. We can, however, determine that there were 5583 entries for this species, and we can proceed to further analysis of the entries specific to this species by using Table 1. Using this table we see that "21" is associated with the Spring Peeper, meaning there were 5583 entries for this species for the whole database.

2.1.3 Basic statistics

In this section we will provide a rough outline to the database's basic statistics, such as the total number of entries for each module. When it comes to the number of entries, this will be described as a "snapshot" because the database is continually receiving new entries and so the number increases continually.

For the total number of entries for a given table, we will use the nrow(x) function in R, which yields the number of rows, and simply substract 1 to account for the header. We can thus deduce that as of 14/04/2015, there are 20971 rows, meaning 20970 entries, in the frogwath_observation table. The R function to get this information, and produce the graph in Figure 3, is as follows:

```
setwd("H://Master's//")
fwo <- read.csv("fw.csv", header=T)
attach(fwo)
nrow(fwo)
```

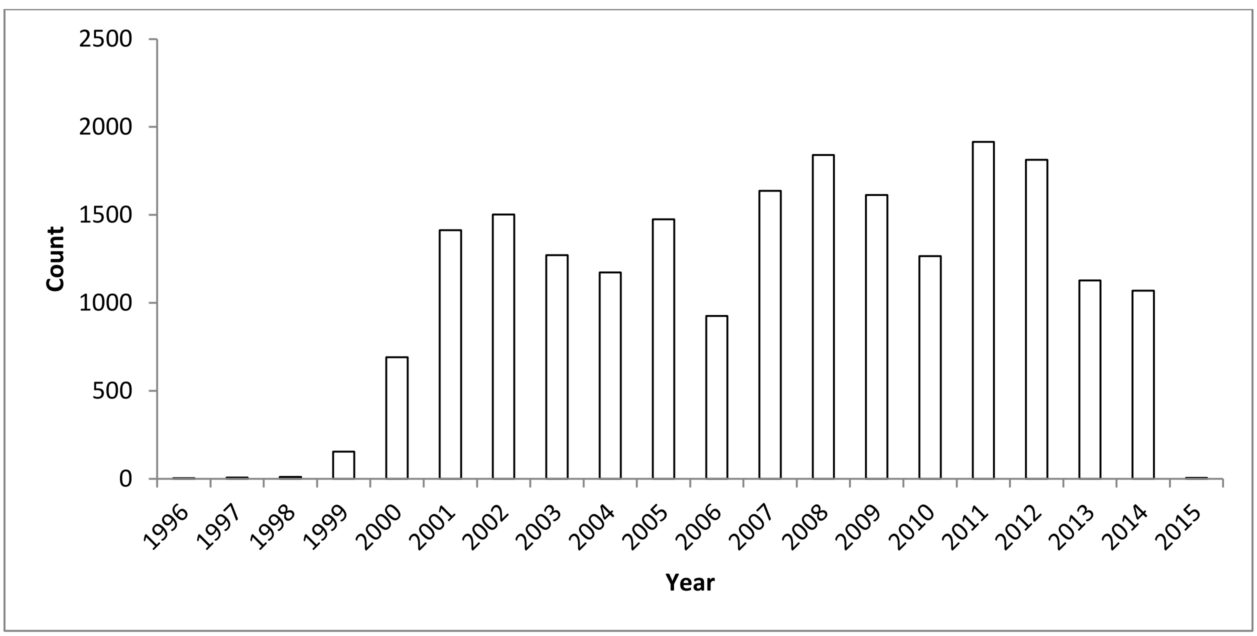

Figure 3: Example histogram of Frogwatch entry count by year.

2.2 Objective 2: Data analysis

2.2.1 Spatial analysis

The spatial analysis is conducted using ArcGIS (version 10.1) in order to map the observations throughout Canada for each module. Maps have also been produced for each species (FrogWatch and Plantwatch) and ice on/off (IceWatch). One benefit of using ArcGIS is its selection tools that allow joint analysis of the spatial and temporal aspects of the data (i.e. where and when a given entry occurred).

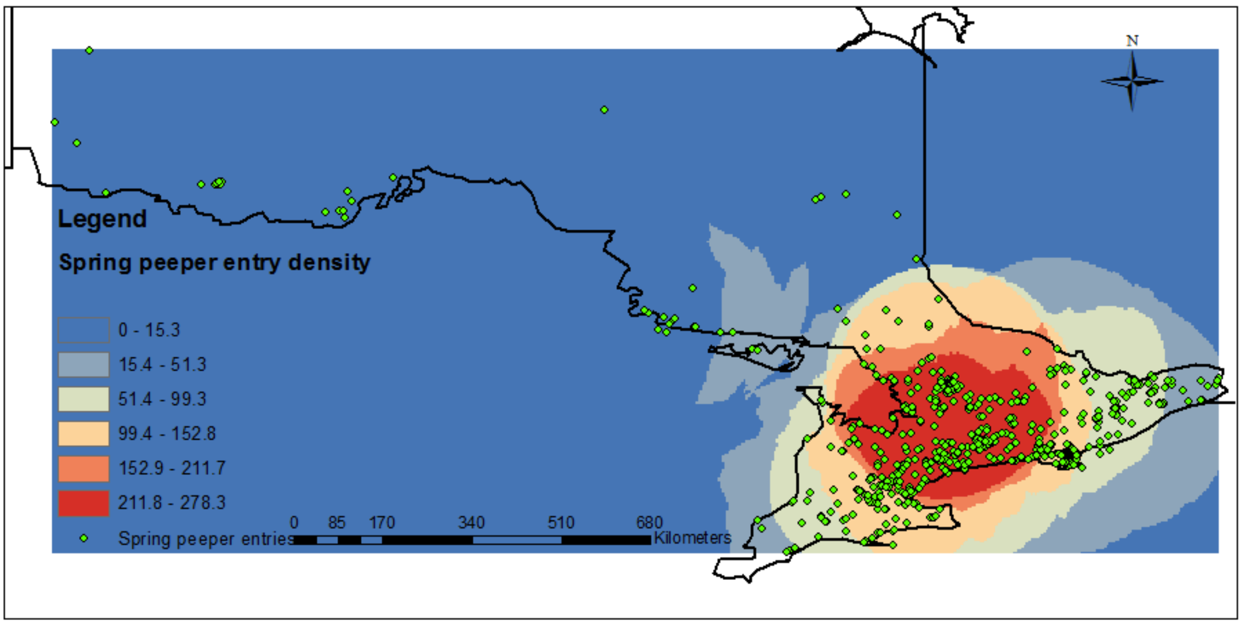

Figure 4: Example mapping of the point density of Spring Peeper entries in Ontario.

2.2.2 Trend analysis

The main softwares that are used for this time-series analysis are R and MS Excel as both of these facilitate time series analyses. Amongst the statistics used include linear regression to explore correlations between variables through time (Figure 6). The annual data was extracted from the date by using the YEAR function in Excel (Table 3). We then used the COUNTIF (meaning counting the number of cells in the provided range which match the given criteria) for each cell adjacent to the new Year column where the range was the entire Year column and the criteria was stated as the adjacent Year column value: this led to every cell in the acount (annual count) column to have the total annual count for the year it corresponded to, meaning we could then do away with the duplicate values to produce a new, much smaller document to run the analysis on with R.

2.3 Objective 3: Example newsletters

The newsletters are presented at the end of chapters 3, 4 and 5. These newsletters are actual scientific analyses on a specific aspect of each individual module (Frogwatch, Plantwatch, Icewatch). These are to be added to the NatureWatch website for reference and to guide future research.

The reason for submitting newsletters connects to the qualities of citizen science and its role as a methodology – being to connect participants with the research. While contributors continue to submit observations, increasing or even maintaining participation requires some form of feedback be made available to the public. In other words, participation requires some indicator that the data gathered is being used for study, hence the production of newsletters.

Chapter 3. Frogwatch analysis

Most of the trend analysis has been done through Excel beginning with a year count analysis to plot the evolution of entry numbers over the program's history. In order to facilitate comparisons between different tables, we subsequently calculated what percentage of the total number of entries occurred each year. In this way it is possible to plot tables with vastly different entry totals (such as the aggregate and a single species), as the values are expressed as fractions of a fixed total (100%). This basic operation also required the elimination of those values without a valid date column entry (and consequently no year value).

While this basic analysis was done for every species of this module, we will only present the graph and map for the aggregate (which will be expressed in count) along with the 3 top species in this section. The remainder (in fact the entirety of the analysis) can instead be found in Appendix A.

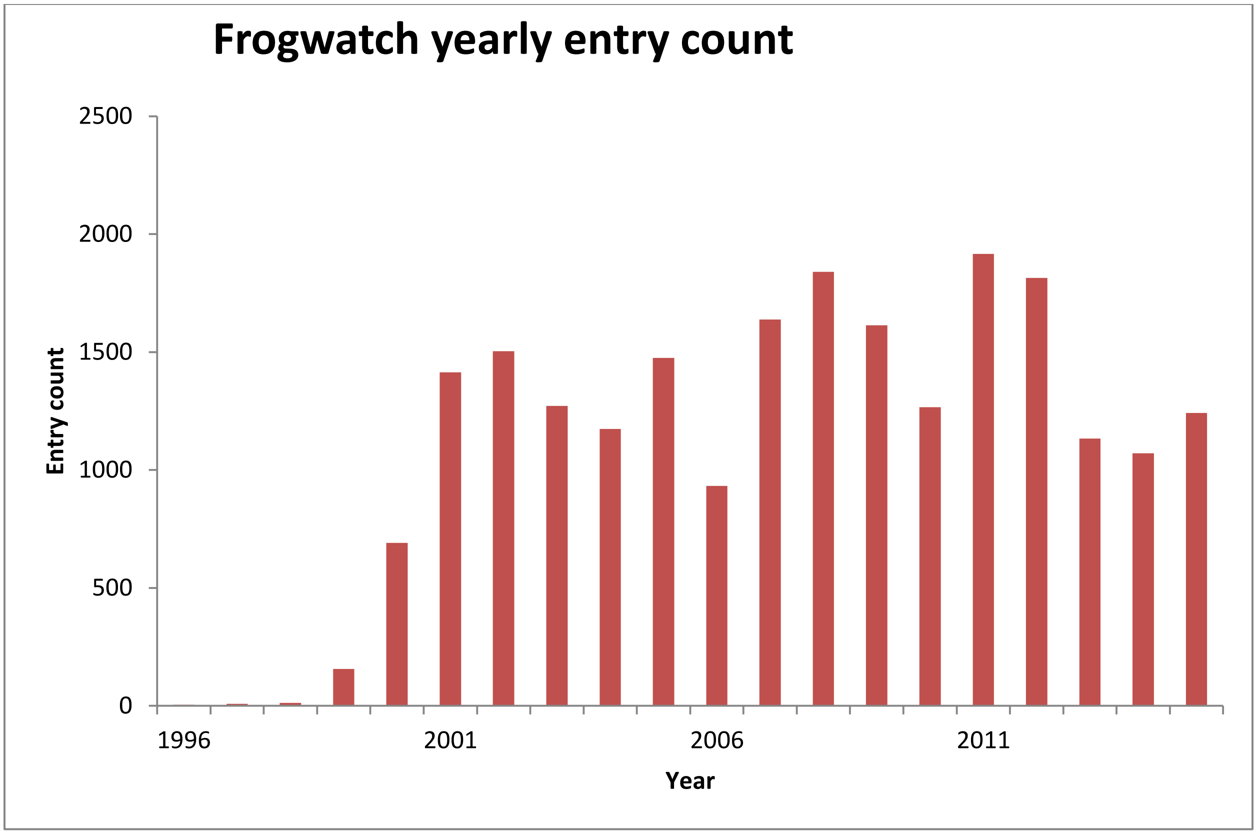

Figure 5: Frogwatch yearly entries from 1996 to 2015.

As figure 5 shows Frogwatch only starts having appreciable entry counts starting in 1999 (the highest before then is 1998 with 12 entries, which is invisible on this graph). We can see a general increase until 2002 where some 1503 entries were sent in, with variations in the following years, most notably a decrease following 2012.

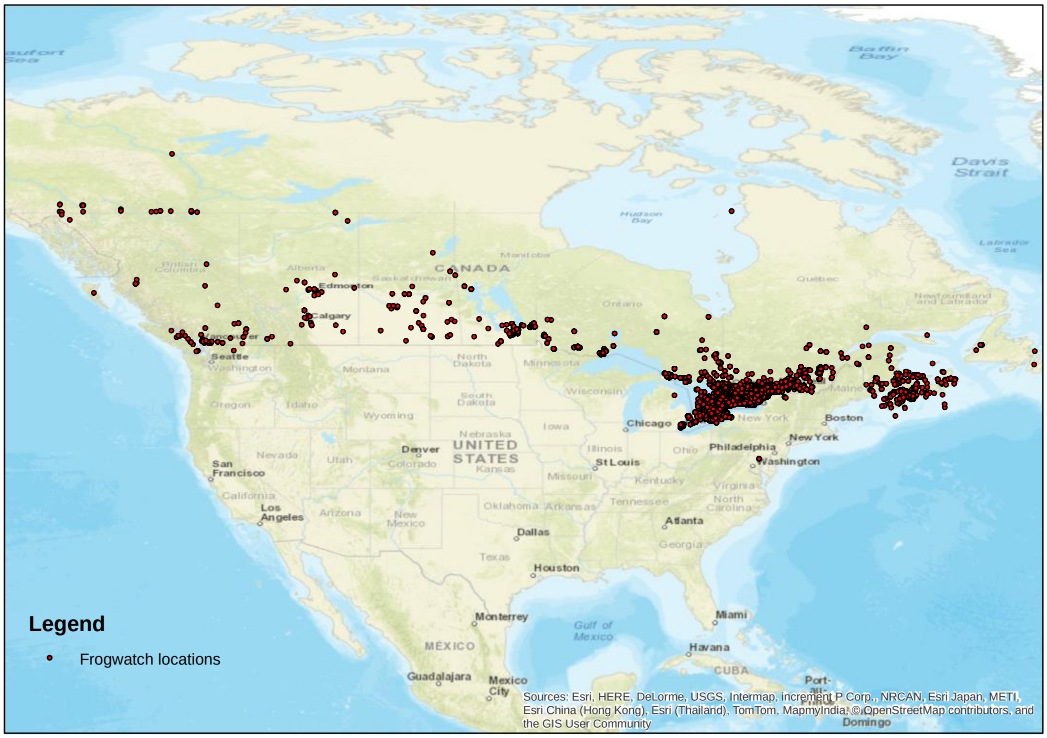

Figure 6: Frogwatch observation locations.

As we can see from the map in Figure 6, besides a small concentration in the Maritimes the vast majority of the entries for Frogwatch occurred in Southern Ontario, around Toronto and near Ottawa.

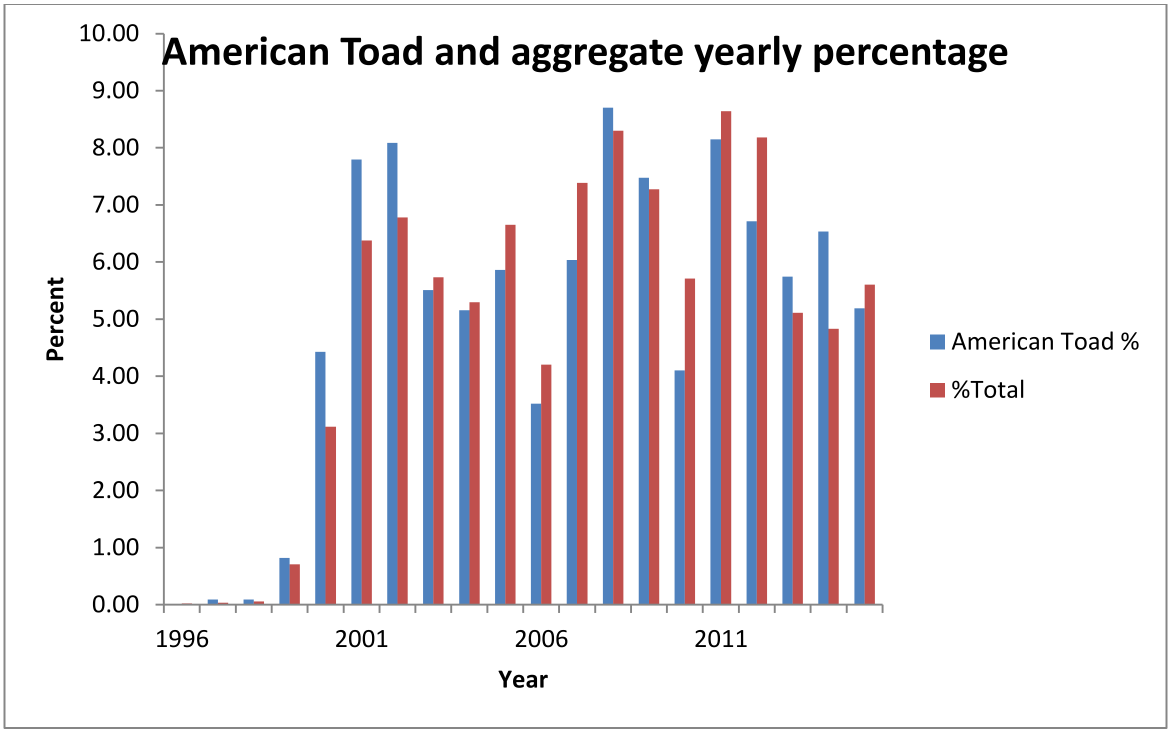

Figure 7: Percentage of entries each year for the American Toad and the total entries.

There are a total of 3412 dated entries for this species. This makes it usable in analysis.

As we can this from Figure 7, the largest difference between the American Toad values and the aggregate is that the 2001 and 2002 peaks are much greater for the American toad (Figure 7). The post-2011 drop in numbers is also less marked than for the aggregate.

We used ArcMap to carry out the spatial analysis (Figure 8). We began by connecting the Locations table and geographically placing them using the longitude and latitude. We then used the join function from the Frogwatch_observation table to the Locations table based on the Location_id column (the other way around prevents repeated entries at the location, which are crucial), and produced a new layer from this join. Following this, we can then select only the entries for each species (using Select by attributes) and produce a layer from the selected features. We now have a point feature layer, and we can use the point density function to produce a raster showing the density per square map unit. We have repeated this operation with every species in this module, and the maps for all the species can be found in the Appendices.

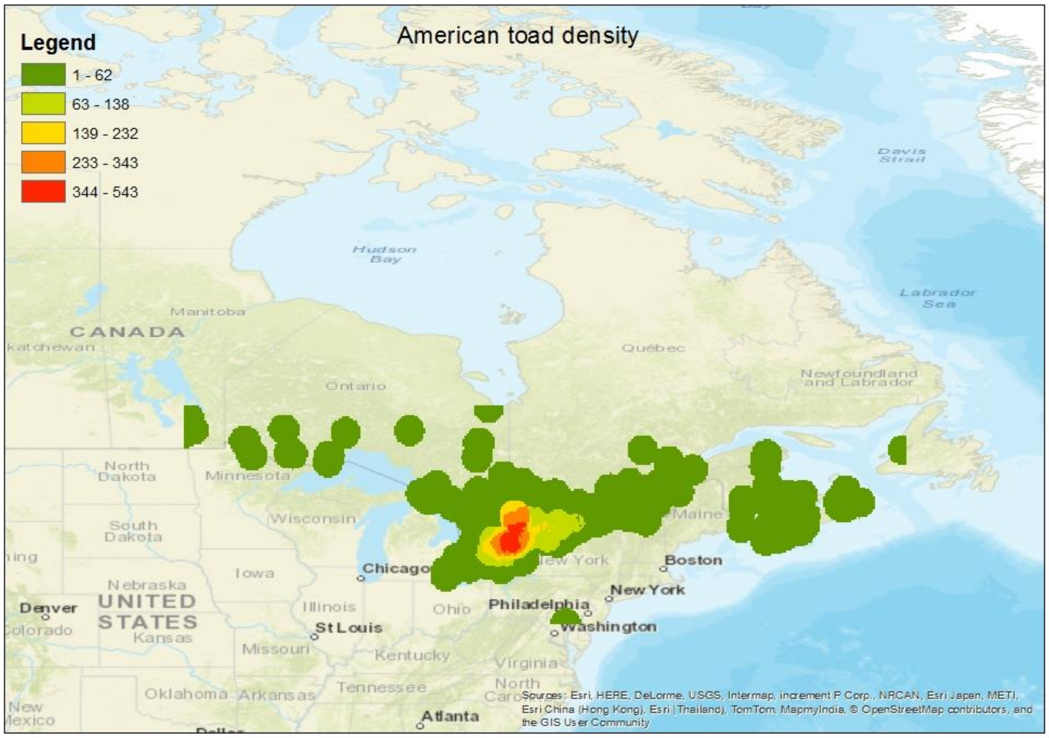

Figure 8: Map of the data point density for American Toad entries.

As can be seen in Figure 8 the vast majority of the entries for this species are located in the Toronto metropolitan area and in the nearby parks.

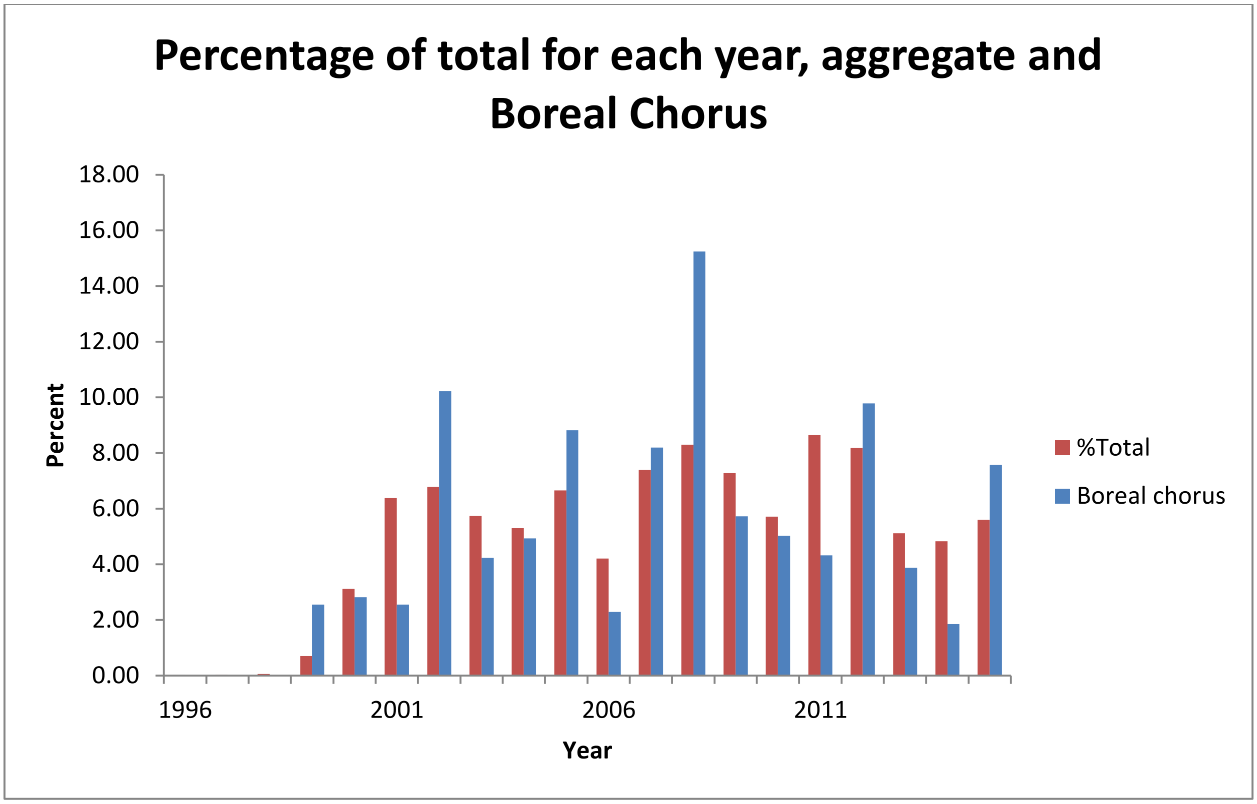

Figure 9: Yearly percentage of the total number of entries for the aggregate and those for the Boreal chorus frog.

There were a total of 1135 entries with usable dates for this species.

We can see in Figure 9 that a number of highly unusual years for this species, notably 2002 and 2008 (high above the aggregate) and 2001 and 2011 (far below the aggregate). In general this species has much greater annual variation than the aggregate, with higher peaks and lower minimum values. This could potentially reflect shifts in populations or the species' calling habits being influenced by weather conditions in a given year.

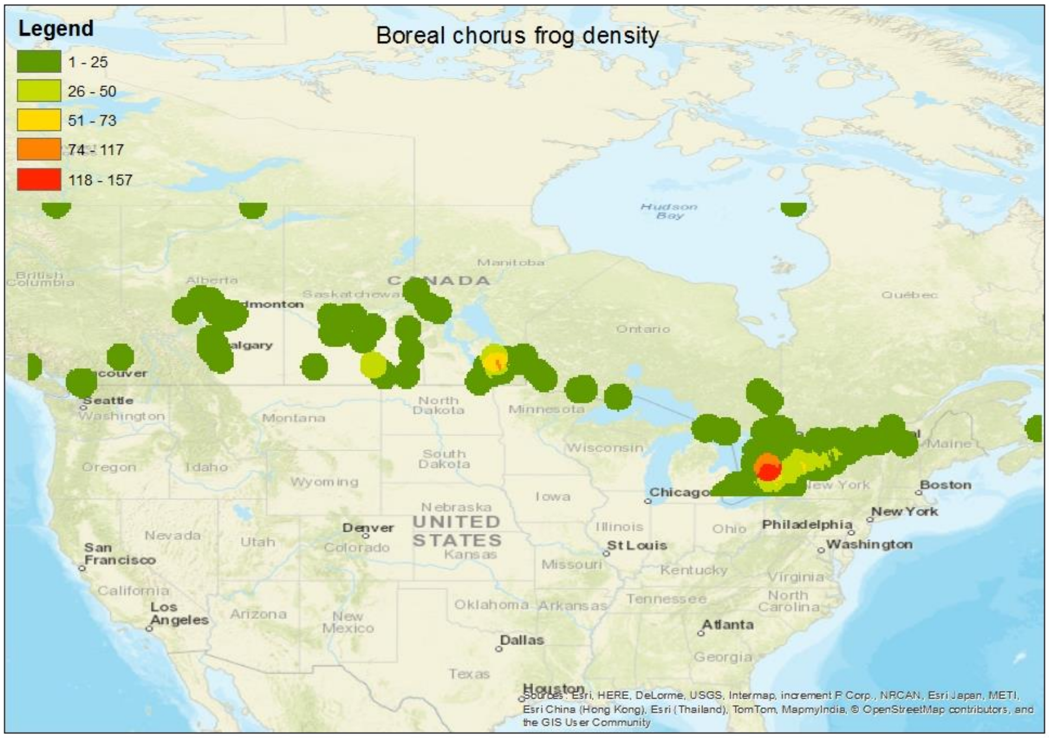

Figure 10: Boreal chorus frog entry density.

Figure 10 shows that there are two main concentrations of Boreal chorus frog entries, namely near Toronto mostly and near Winnipeg as well.

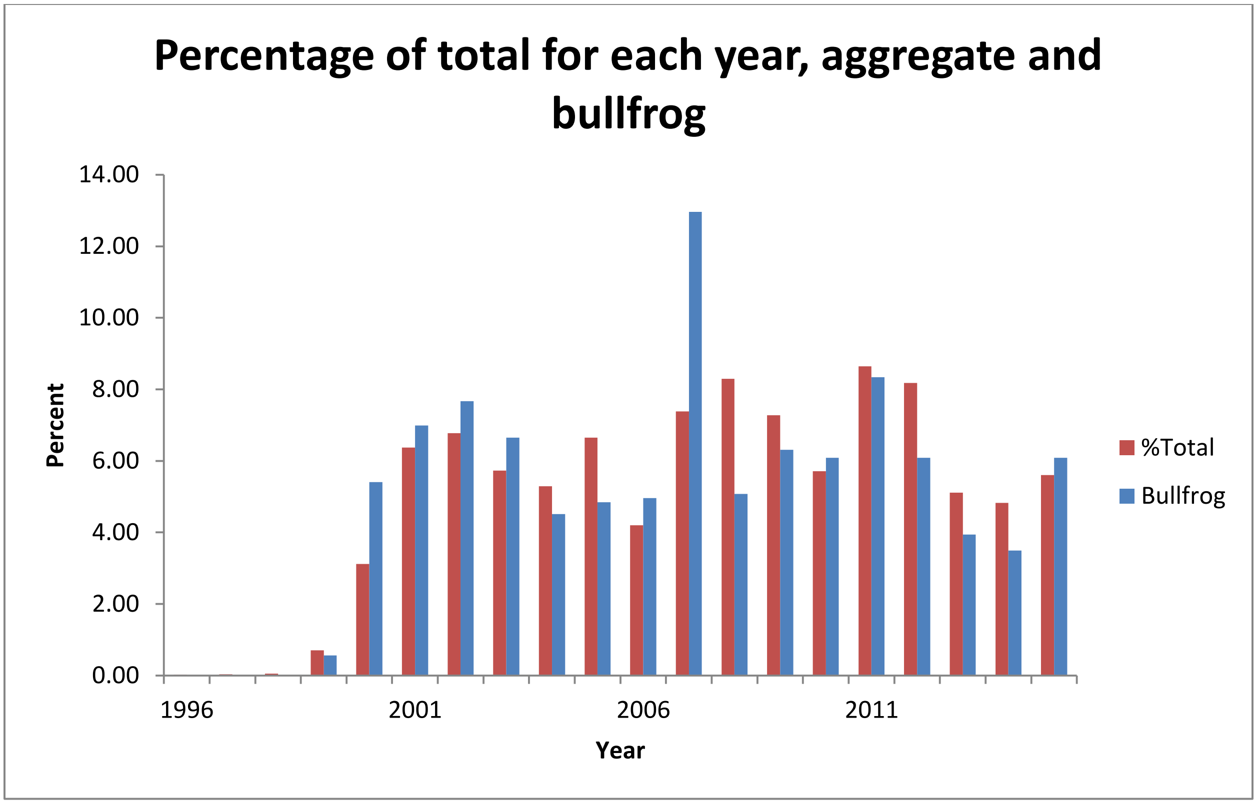

Figure 11: Yearly percentage of Bullfrog entries and Aggregate.

The total number of usable entries for this species was 887.

The Bullfrog entry numbers largely matches up to the aggregate as shown in Figure 11, except for 2007 where the proportion is much higher. Considering how great this difference is and how every other year closely follows the aggregate, investigating the conditions and events of 2007 would be of particular interest for future research.

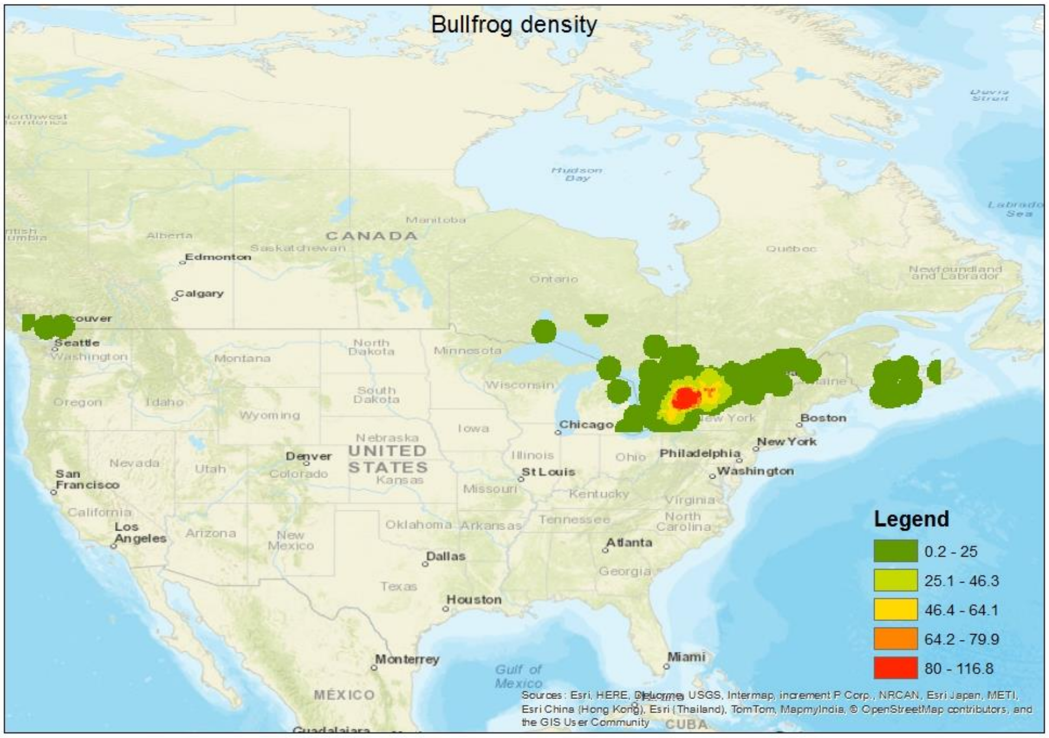

Figure 12: Map representing the entry density for bullfrog data points.

Most of the entries for this species were concentrated around the Toronto area as Figure 12 shows, in common with many other Frogwatch species, which matches up to expectations for a Citizen science program having entries concentrated in and around human population centers.

3.1 Frogwatch newsletter

Figure 13: Leopard frog.

Figure 14: Spring peeper.

The evolution of the number of Northern Leopard Frogs (Figure 13) in Canada is a subject of some concern. While fairly widespread this species does seem to be in decline in many areas (Naturewatch, 2016). As a result of this situation we will seek to try to determine, from the NatureWatch data, if there is indeed such a decrease.

Note that we will continue to use Leopard Frog to refer to the Northern Leopard Frog throughout, as it is the only species of this family in Canada.

Why 2 species?

In cases where monitoring the environmental conditions over time is difficult, we can turn to the analysis of several different species with varying living requirements to infer these conditions based on how these species' numbers change over time. In effect this is like estimating how a stream bed is shaped by looking at how the flowing water swirls and eddies.

With this in mind, we can see how a database such as NatureWatch can be used for basic analyses – by monitoring such a large number of species over a very large area, it can allow for comparisons between species known to be of "least concern" (as a baseline) and species whose status is more compromised. We can also extract the numbers for every species monitored to produce an aggregate value, which can be used to partially isolate the effect of participation on the entry counts, which is an important consideration for citizen science.

For a baseline comparison we will use the Spring Peeper (Figure 14), which is the species with the single greatest number of entries in the database at well over 6000. This species is to be "stable throughout most of its range" (Naturewatch, 2016) – we will see if the results indicate a similar conclusion when compared with the aggregate.

The sample size of the Leopard Frog is also quite large enough for analysis as well with over 2000 entries.

Why southern Ontario?

We have decided to focus on analysing weather archives for Ontario, more specifically southern Ontario, as this is where the vast majority of the entries for the two selected species occur.

Given the fact that both Leopard Frogs and Spring Peepers largely occupy similar areas in the same regions we can surmise that within this region both species experience the same weather.

Numbers and trends

From the data in Figure 15, we can see that the Spring Peeper's yearly percentages roughly match the aggregate and do not show a particular decline relative to the aggregate.

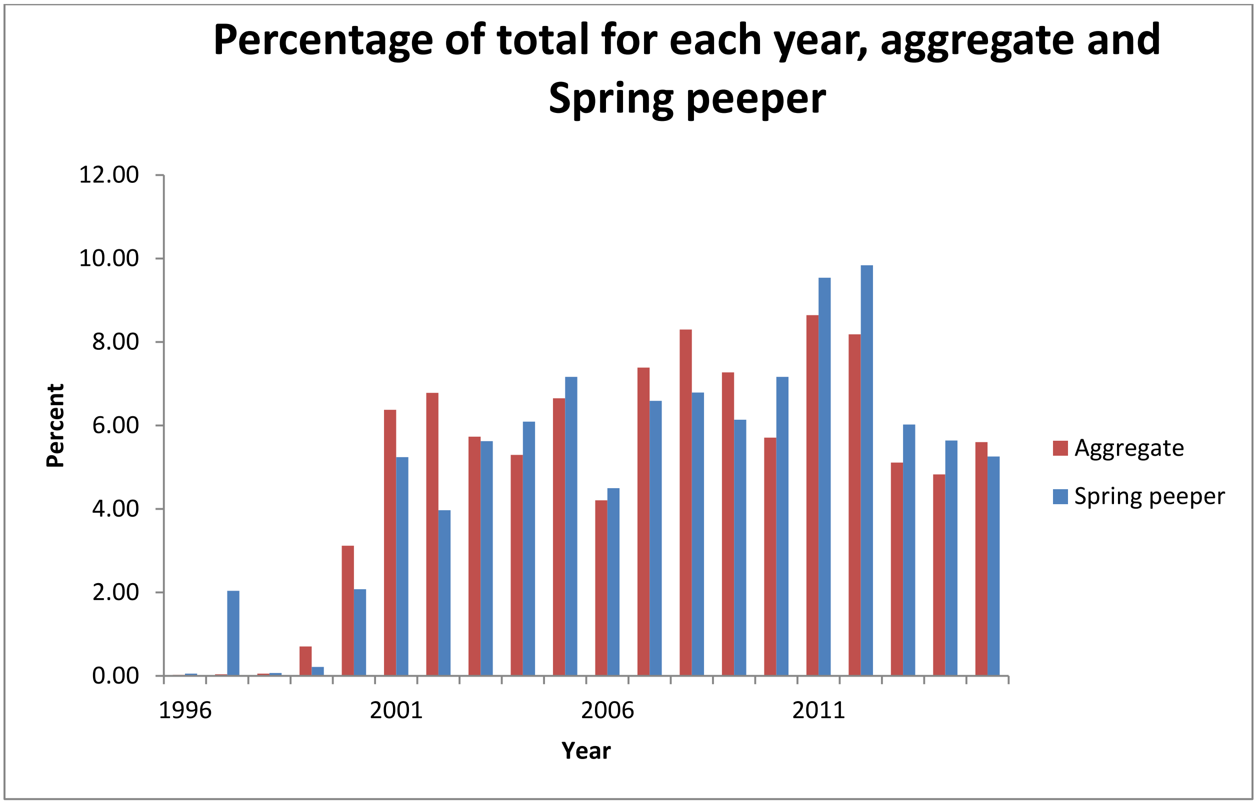

Figure 15: Yearly percentage for the Spring peeper and aggregate datasets.

With this in mind, we can now look at the data for Leopard frogs and establish whether and when major divergences may have occurred.

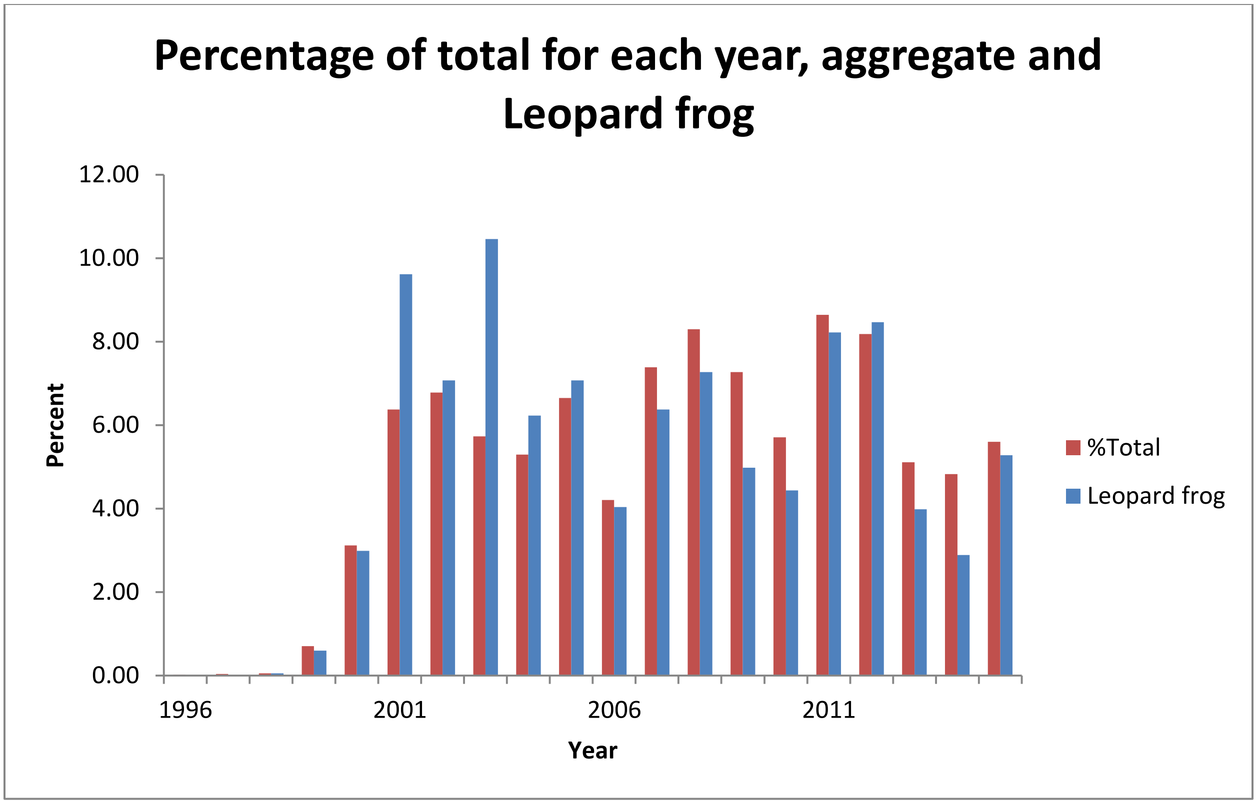

Figure 16: Yearly percentage of total entries for the Leopard frog and aggregate from 1996 to 2015.

As we can see from Figure 16, the Leopard Frog data indeed has an inverse relationship with the aggregate which seems to accentuate over time. However what is most obvious is that the perceived decrease seems to mostly be a result of the unusual entry numbers of Leopard Frogs in 2001 and 2003.

If we consult the NatureWatch page for Leopard frogs, we find that droughts are a factor in their population (namely because of their requirement for permanent ponds for reproduction). It also states that the species attains maturity at two years of age. That the peak years (relative to the aggregate) are two years apart, they may be the result of a single strong signal, with those born as a result of the 2001 boom leading to an abundance of reproduction in 2003, when they have attained sexual maturity. However it is also possible that both of these years reflect conditions in their respective year.

The weather

Weather data archives indicate, from a cursory observation, that the summer of 2001 was one of prolonged drought in southern Ontario (ECCC, 2016). This does not seem conducive to the Leopard frog at first glance, but this belies other important facts when we wish to consider water bodies. The ECCC page for the Top Ten Weather Stories of 2001 mentions that this year saw a rise in the level of the Great Lakes due rainfall and runoff in the spring (ECCC, 2016). This may go some way towards explaining why the Leopard frog population may have done so well in this year, with permanent ponds bolstered in the spring to such an extent that they survived despite the potent drought that followed in the summer.

If the cause for the 2001 spike in Leopard frog observations is due to environmental conditions being favorable for their survival and reproduction, we ought to expect a similar pattern for other species with similar requirements living in the same region. We do see such a peak in 2001 for the Green Frog as well, which depends on permanent ponds on a daily basis and also for reproduction.

Conversely if this hypothesis is correct species in the same area with different or simply less strict requirements would register no such correlation or even an inverse correlation. The Spring Peeper certainly seems to respond in such a way, with slightly lower values for 2001 than the aggregate. The website asserts that this species can reproduce in a wide variety of habitats, notably temporary woodland ponds. Given how wet the spring was in and how dry the summer was many such temporary ponds would have vanished possibly (however the difference with the aggregate is not strong).

This analysis allows us to assert that the citizen science data collected through this database can be a useful tool for basic analyses, and for monitoring species over a large geographic scale.

Chapter 4. Plantwatch analysis

The analysis done for Plantwatch is very similar in methodology to that of Frogwatch. As both datasets have a "species" value, plotting similar yearly entry counts for each species is the main port of call for this general analysis. The same applies for the spatial analysis for this module.

Also in common with Frogwatch, the graph descriptive paragraphs will always be referring to the percentage when comparing the species to the aggregate even if it is not specified.

The primary difference is of course the monitoring method, which is to look for specimens in their environment. Unlike with Frogwatch where the observations are mostly made by staking out in a familiar area and listening for calls during the twilight, the Plantwatch module observations are mostly done during the day.

The three species selected for this are Aspen Poplar, Saskatoon and Dandelions. This is because all three of these species had very high entry counts overall, meaning that as a sample they were quite suitable thanks to the size of said sample.

The brief analysis for all the species (meaning the graph and map) can be found in Appendix B.

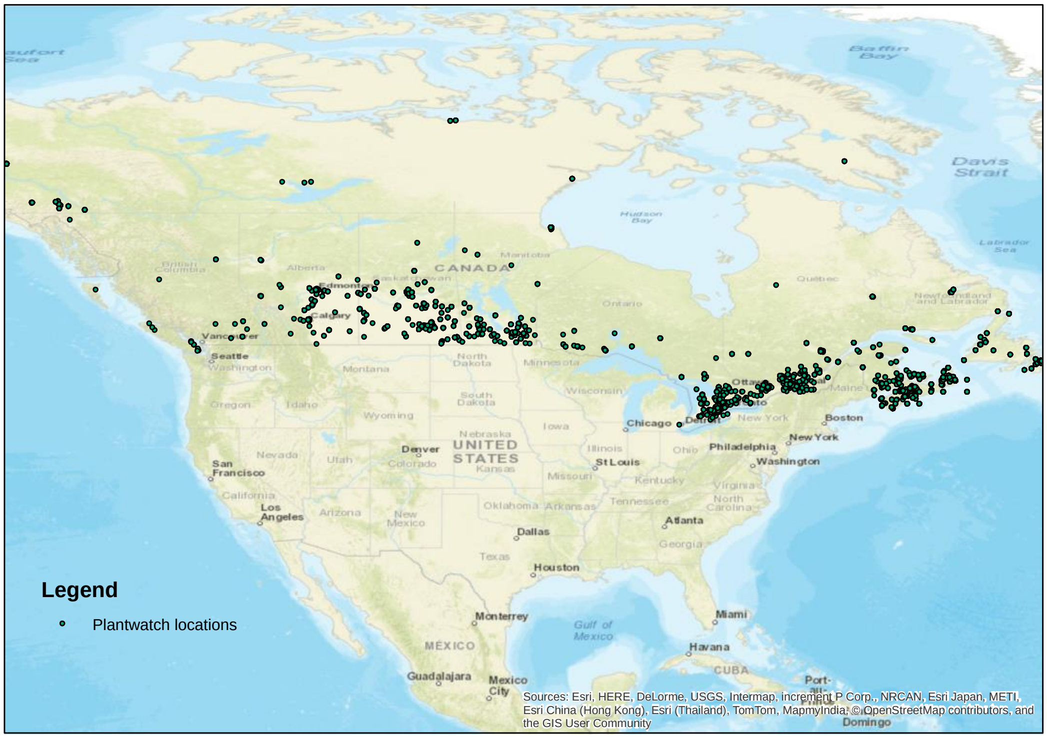

Figure 17: Map of all locations of Plantwatch events.

Figure 17 shows the distribution of sites for Plantwatch. Clusters of Plantwatch entries appear in Southern Ontario (Toronto), in the Ottawa region, the Atlantic region and the Prairies.

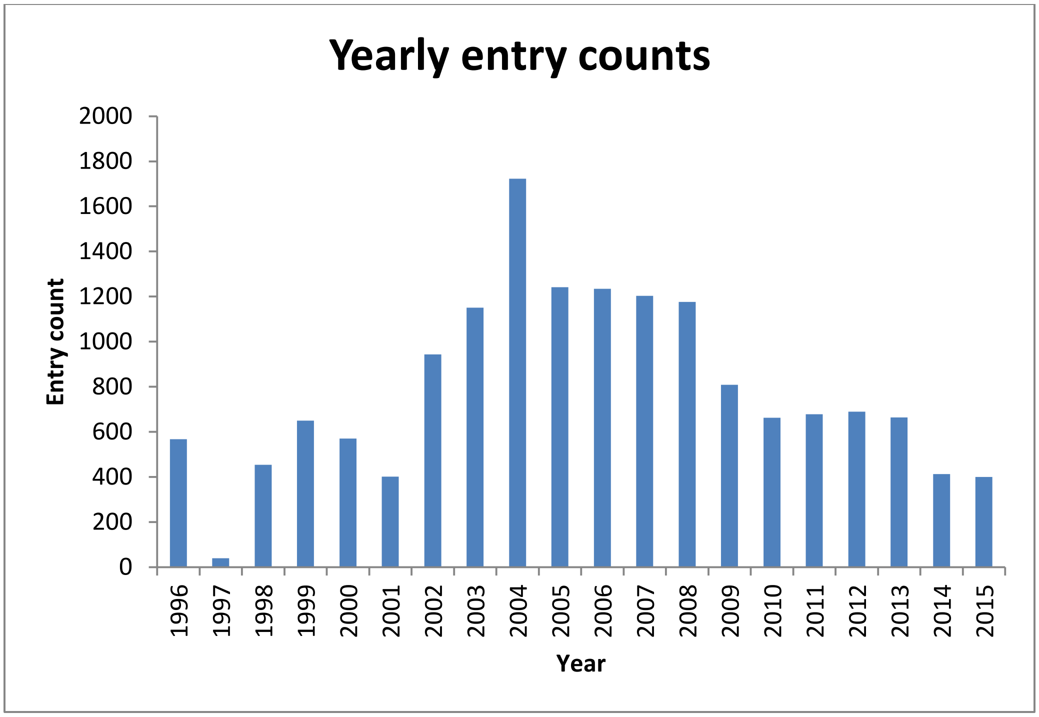

Figure 18: Total number of Plantwatch entries for each year from 1996 to 2015.

Figure 18 shows the yearly total number of entries only from 1996 to 2015. The actual data goes as far back as 1956, however before 1996 the highest number of entries is 14, in 1980 and every other year before 1996 is either in a similar order of magnitude or simply nil. In either case the numbers were far too small for these years to allow any significant analysis.

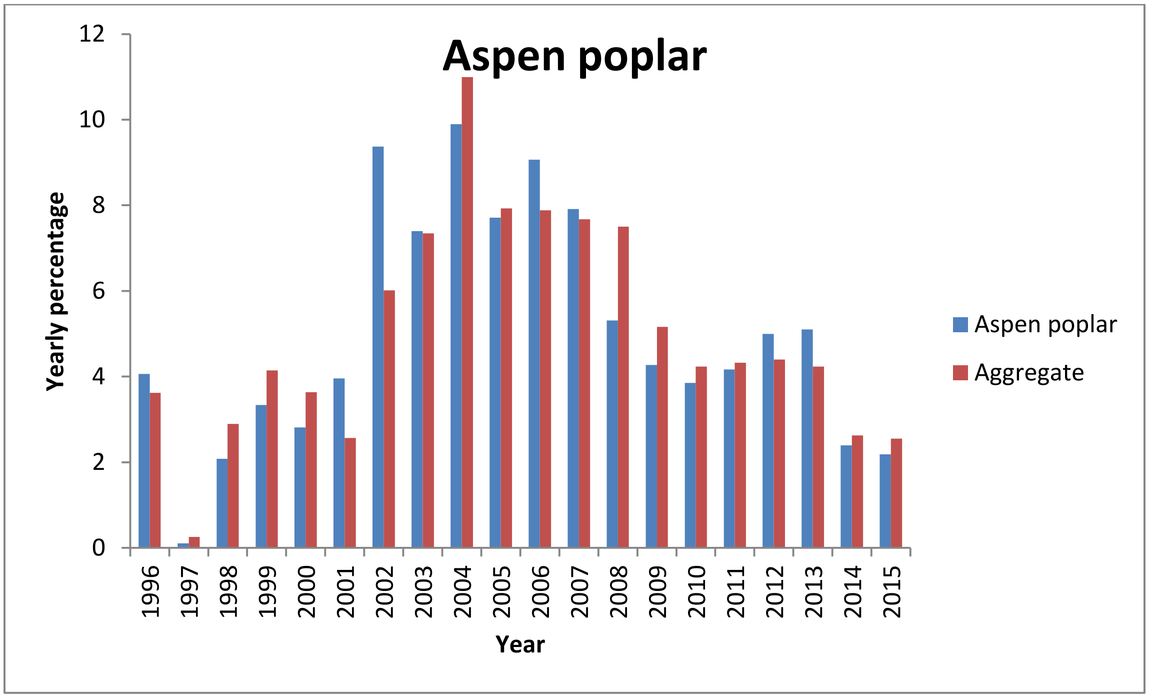

Figure 19: Yearly percentage of the total number of entries for the aggregate and those for Aspen poplar.

The number of entries for this species, being 983, means it has fairly good value for analyses. As Figure 19 shows it mostly follows the aggregate percentage values, except for 2002 where it rises above the aggregate. This species may be of interest for analysis as a result of its overall consistency except for this year.

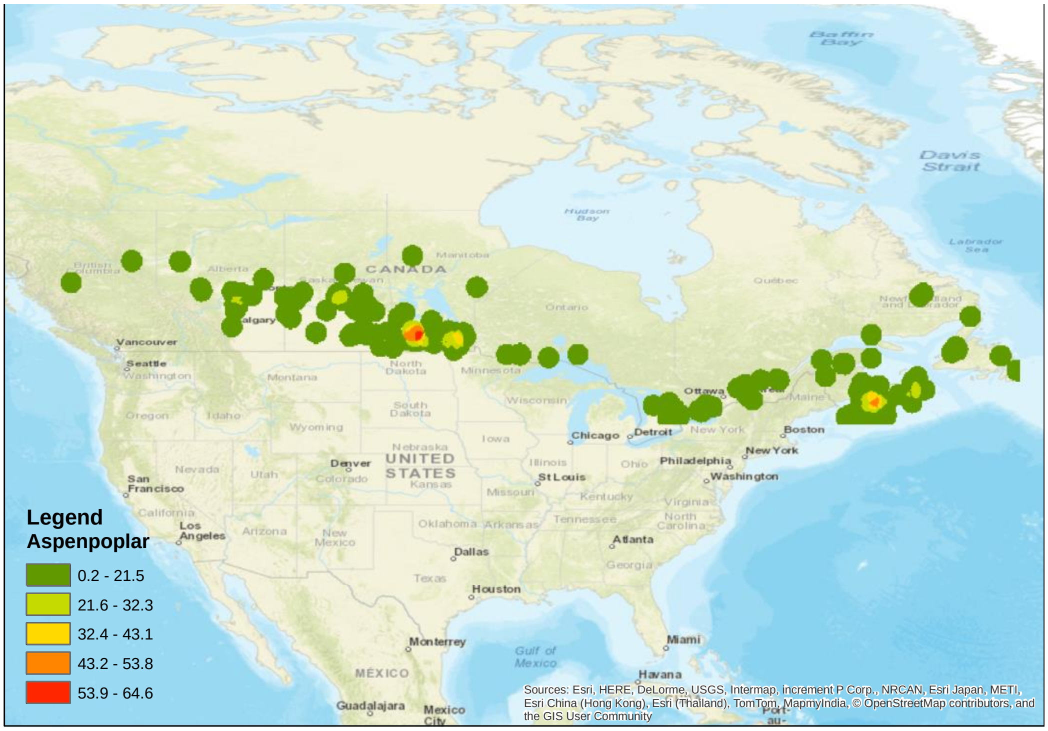

Figure 20: Map of the area density of Aspen Poplar entries.

This species is atypical in that most of its entries are concentrated to the West of Lake Manitoba and at the borders of New Brunswick, Prince-Edward Island and Nova Scotia as Figure 20 shows rather than in the Toronto area.

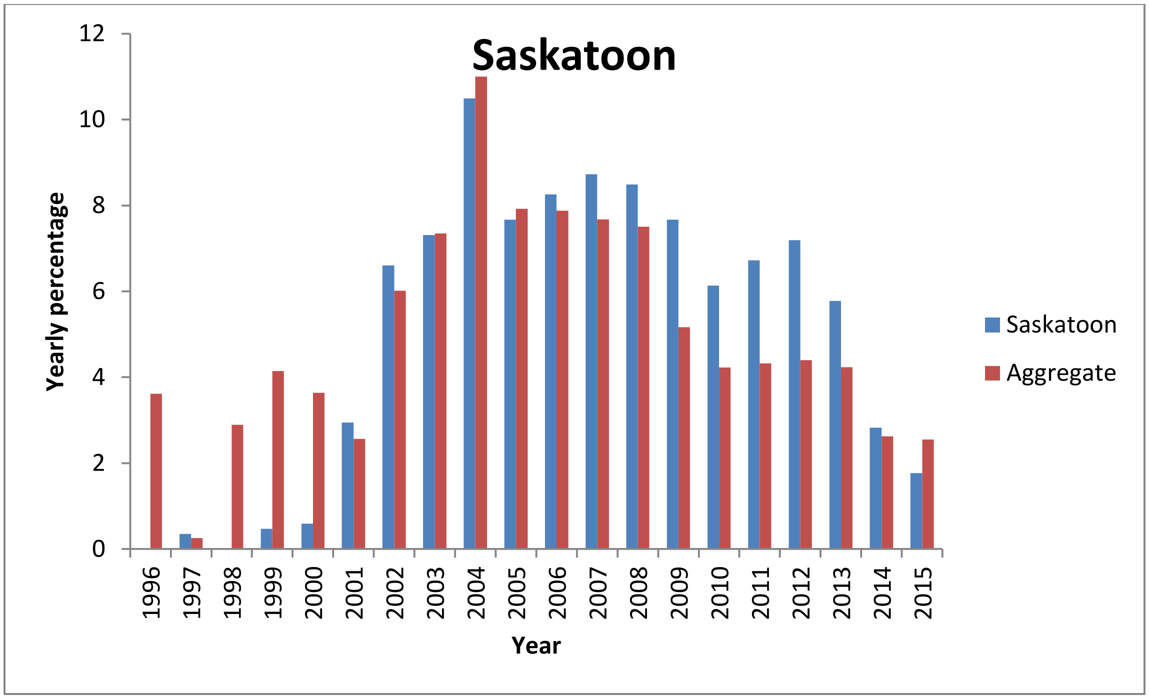

Figure 21: Yearly percentage of the total number of entries for the aggregate and those for Saskatoon.

The Saskatoon is another good species for analysis with 877 entries. As Figure 21 shows this species is notable for having percentages noticeably above the aggregate from 2009 to 2013, otherwise it follows the aggregate from 2001 onwards (with few entries at all before this year). The consistency of the yearly percentages for this species after 2001 makes it interesting and may indicate it would be useful as a control for certain analyses.

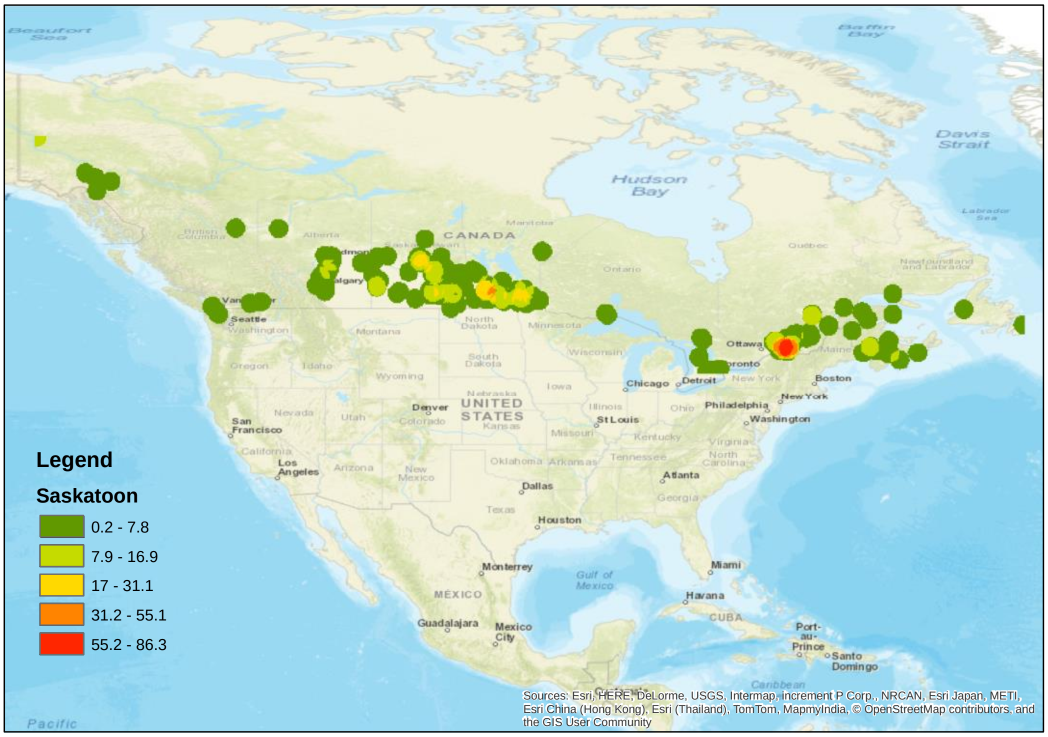

Figure 22: Saskatoon entry density map.

As per Figure 22, we see that the entries for this species are concentrated near Montreal and to the West of Lake Manitoba, with barely any in the Toronto area. This is something of an oddity when the most entry-dense area of the module lies between the two main concentrations for a species but has almost none. Investigating why this is would be an interesting question to answer, as the two hubs for this species are in totally different environments.

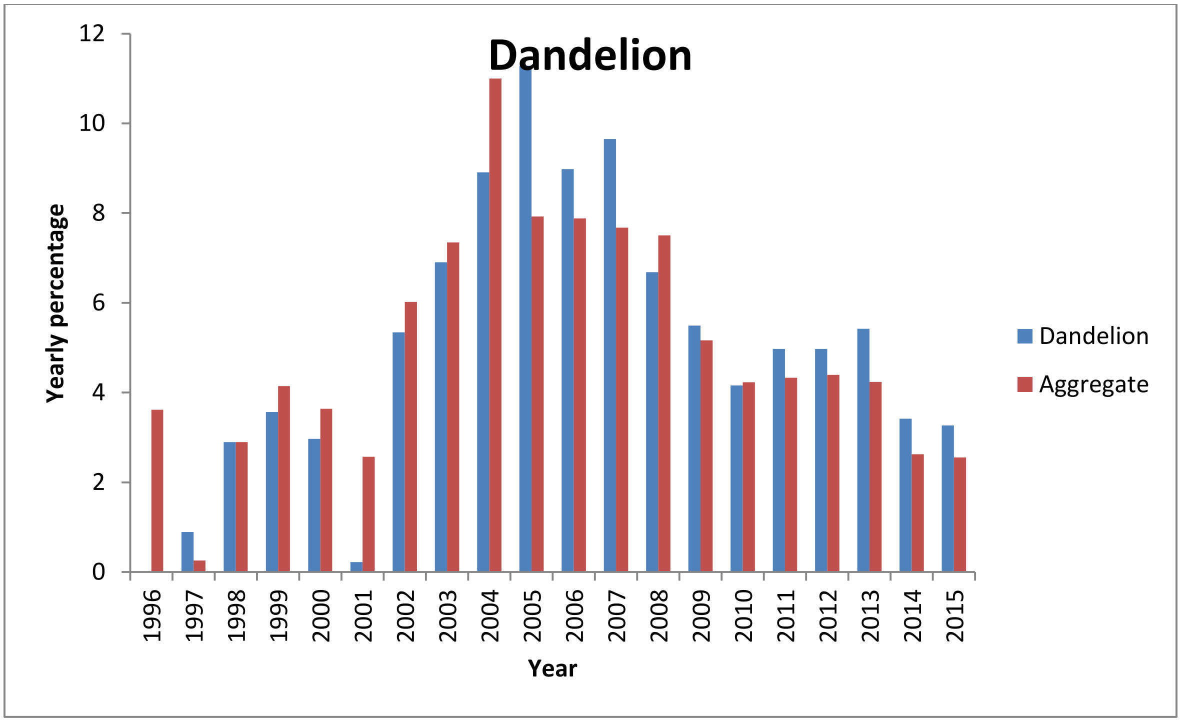

Figure 23: Yearly percentage of the total number of entries for the aggregate and those for Dandelion.

This introduced species is extremely abundant (with over 1366 entries before 2016). It also generally follows the aggregate very closely except for 2004 and 2005, as we can see from Figure 23. Given how widespread it is and how it thrives in human-made lawns unless managed it is not the best species for analysis though, except perhaps as a means to isolate the entry counts for other species from participation (as it is so easy to find).

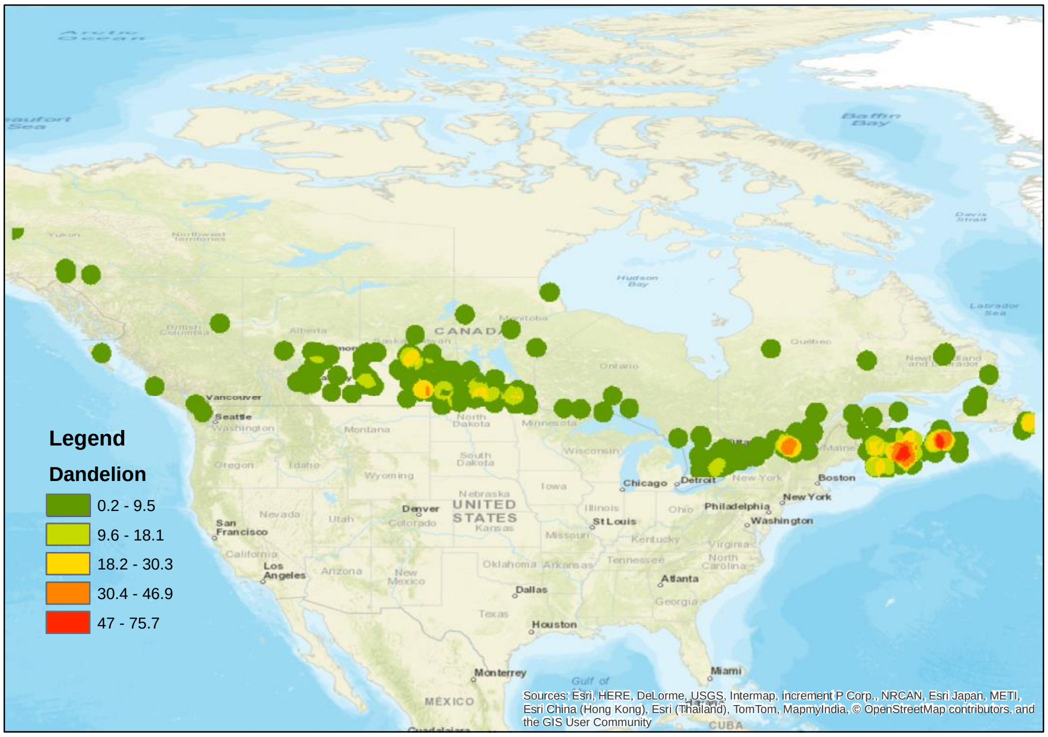

Figure 24: Dandelion entry density over Canada.

As shown in figure 24, Dandelions are a very widely observed species, with many different areas of relatively high concentration, on Prince Edward Island, between Ottawa and Montreal, and in several areas of the prairies provinces.

4.1 Plantwatch newsletter

<u>Plantwatch</u>

Using the Plantwatch data for research requires a slightly different approach given the very distinct (and in fact much more diverse) species being analysed. While frog populations (or at least noticeable populations) can be correlated to recent or current conditions given the fairly quick population response and the mobility of these species, the same does not hold true for plants.

Obviously on an individual level mobility is a non-issue with plants, even if populations can effectively migrate over successive breeding cycles, but the species in Plantwatch show much more variation in life history than those of Frogwatch. While some species survive over the winter year after year, others will die off leaving their place to the next generation every year. Even among the perennial plants there is a large difference in how "lasting" they are. Certain tree species have individuals who have been alive longer than the oldest citizen science program in North America (the Audubon Christmas bird count) – to say nothing of Naturewatch, which is barely 15 years old. While this obviously makes it very difficult to monitor changes to such long-lived species over time, it does make them excellent to isolate participation as a variable. This measure relies on the fact that the trees are, for the purposes of the database as of yet, largely unchanging (which is an oversimplification but can provide a means to control for observer numbers).

Top species - controls

We can also use shorter-lived but ubiquitous species which tend to thrive near human settlements to further control for participation. Chief among these species is the Dandelion (Figure 25), the well-known flowering weed imported from Europe, so common on private and public lawns that a few gardening tools are used almost exclusively for the removal of these taproot-bearing plants. We can make use of the fact that this species is overabundant to the point of being considered a weed and reproduces vigorously in the warm seasons with far-spreading wind-borne seeds. When it comes to weather and environmental parameters dandelions, true to form for an invasive non-native plant, is incredibly hardy: it can tolerate partial shade, is excluded from only the most extreme soils (flooded, completely dry sand) and is a true perennial regardless of winter conditions. It could be argued that it is quite resistant to most pests as it can regenerate completely (and repeatedly) from its taproot – indeed completely removing the plant in its entirety is the only way to guarantee it does not recover (even then if done improperly that can cause multiple smaller shoots to grow from the spot).

Figure 25: Dandelion flower in bloom.

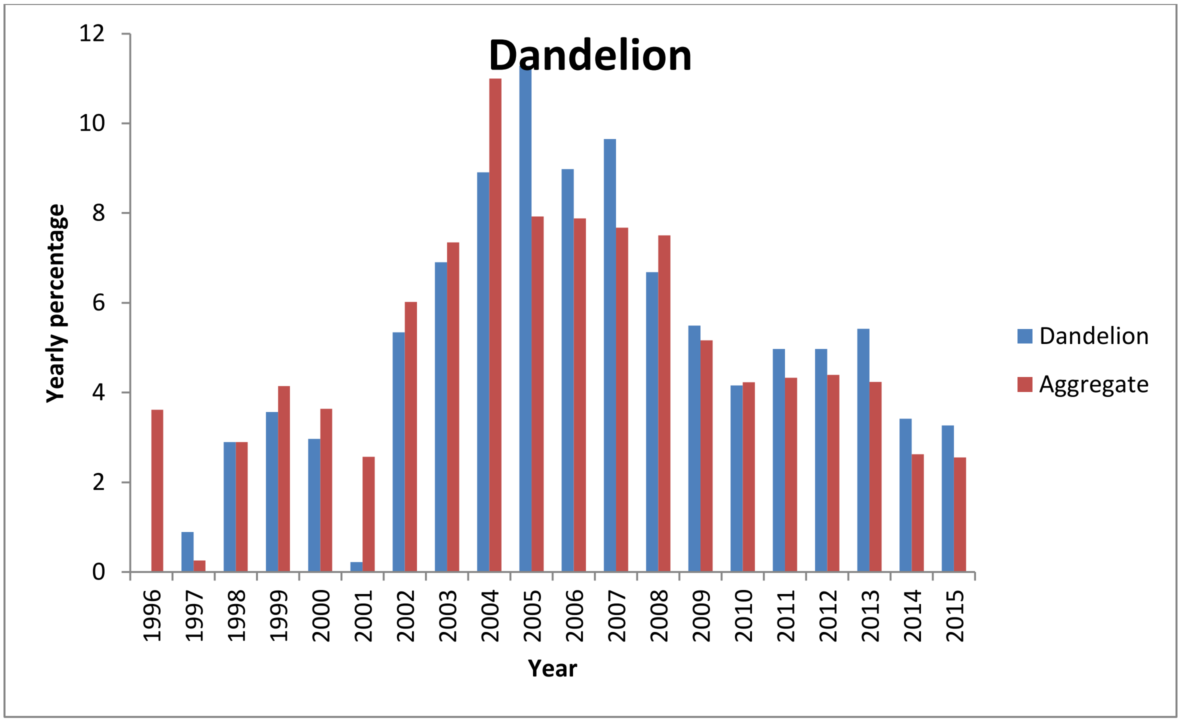

Figure 26: Dandelion yearly percentage of total entries compared to the aggregate from 1996 to 2015.

As figure 26 shows, the dandelion is fairly consistent in following the aggregate after 2001, the year when Naturewatch was instated. This is in exception to 2004 and 2005, as the aggregate peaks in 2004 and dandelion peaks in 2005. These observations are to be expected for such a ubiquitous and hardy species.

We will also consider the Common purple lilac (Figure 27), as this non-native species is mostly planted in lawns on property – meaning its prevalence is largely insulated from climate conditions, and that it is quite easy to find close to where observers may live. It is also a relatively long-lived species, meaning it is not subject to much yearly variation.

Figure 27: Common purple lilac blooming flowers.

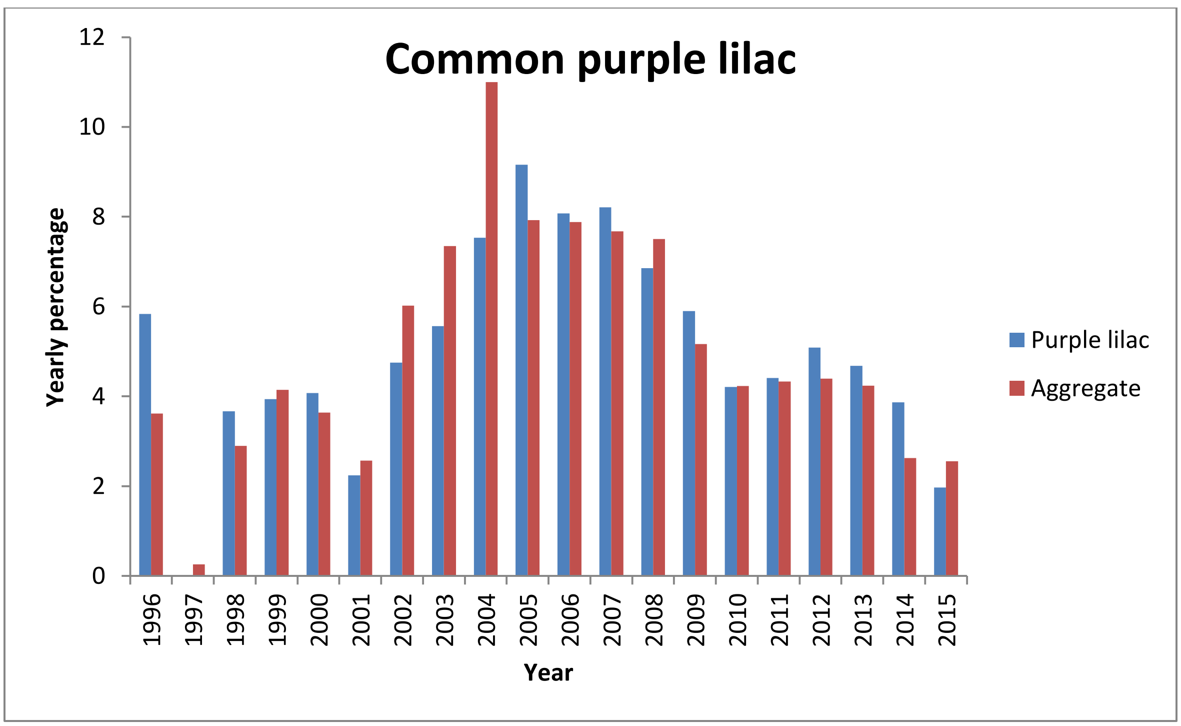

Figure 28: Common purple lilac and aggregate yearly percentage of total entries, from 1996 to 2015.

Figure 28 shows that of the Common purple lilac is one of the species which is closest to the aggregate in terms of trends other than for 2004.

For comparison purposes we are also considering the Wild Strawberry (Figure 29). This plant has low shade tolerance but is otherwise hardy.

Figure 29: Wild strawberries bearing fruit.

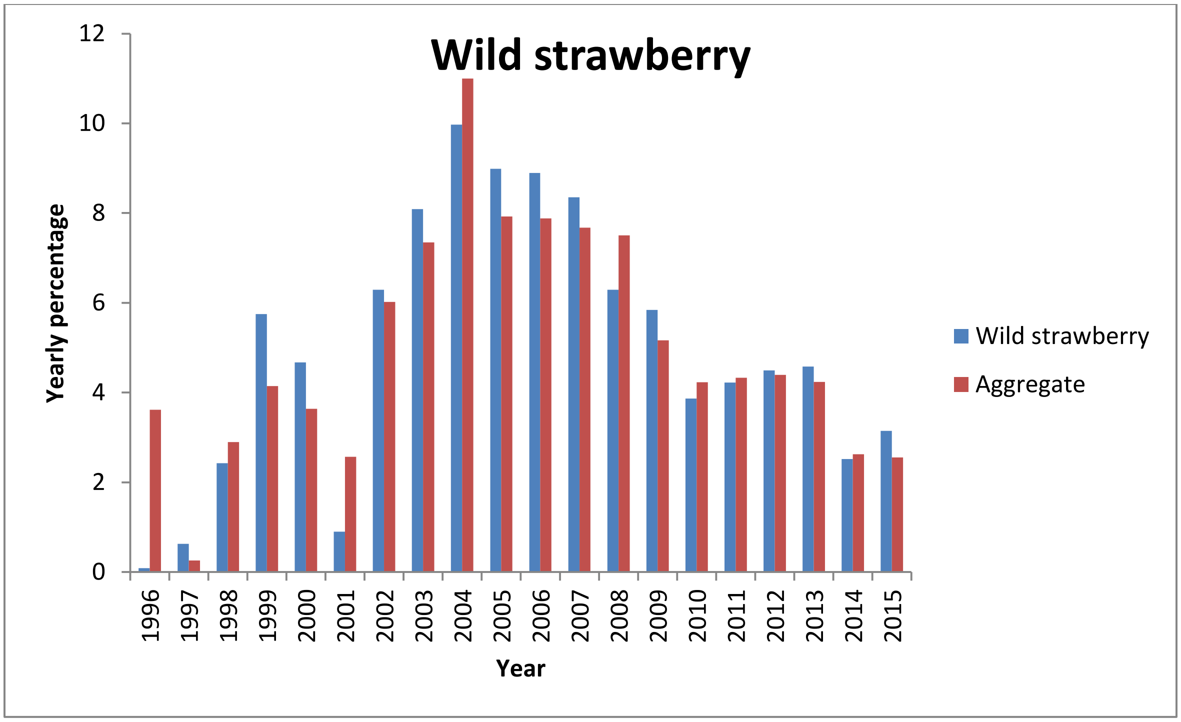

Figure 30: Yearly percentage of the total entries for Wild strawberries and the aggregate from 1996 to 2015.

Figure 30 shows how closely the Wild strawberry follows the aggregate's trends, except for 1996. This would hint at this highly sought-after species (due to its edible berries) having good value as a baseline.

As we can see from these descriptions, the species chosen to serve as baselines are known for their hardiness, which in many ways isolates their populations from dramatic changes in the seasonal cycles.

Why Aspen Poplar?

We have selected the Aspen poplar (Figure 31) for analysis because this species has a substantial sample size and is fairly widespread.

Figure 31: Aspen polar flowers in the early bloom stage.

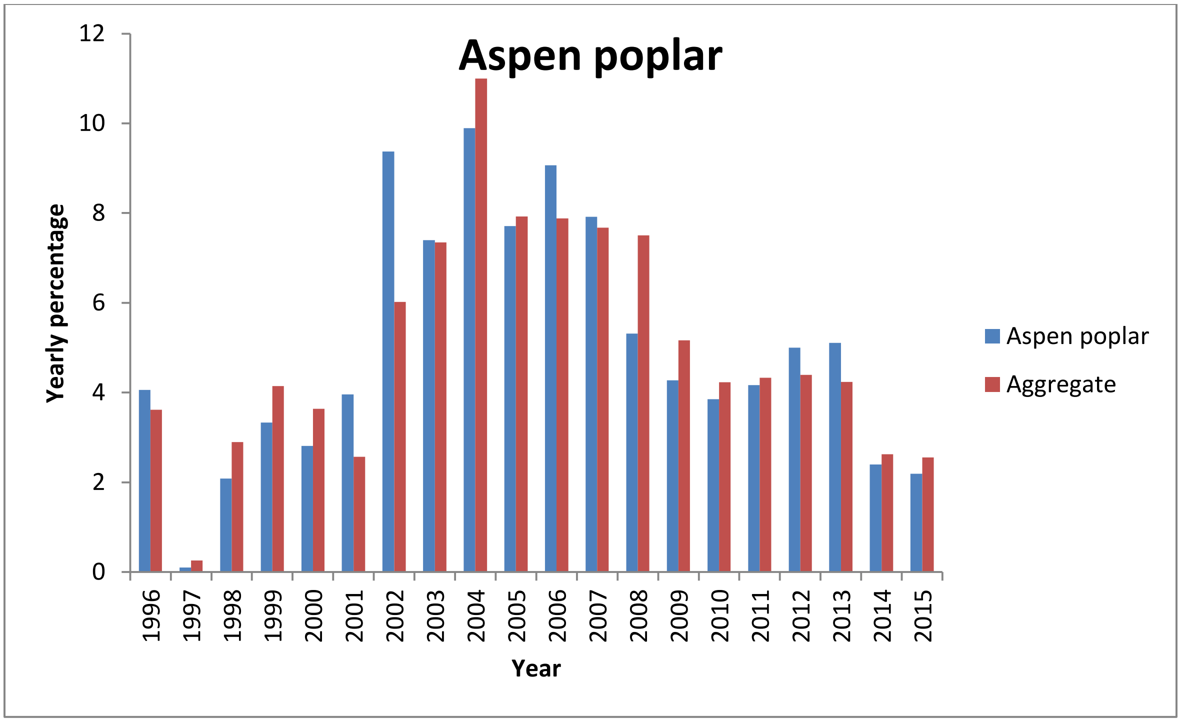

Figure 32: Yearly entry percentage for Aspen poplar and the aggregate from 1996 to 2015.

Figure 32 shows that the Aspen poplar data has an unusual peak in 2002. Considering that none of the baseline species used had a particularly high number of entries for that year this fact may indicate truly unusual conditions for 2002.

It should be noted that Plantwatch instructions focus on male Aspen Poplar individuals (Naturewatch, 2016).

The time scale of Aspen poplar's life cycle, namely with regards to its seeding and regeneration, make interpreting the results difficult. Currently the methodology does not allow for the differentiation between observations made of mature trees and relatively new growth (Appendix B). Given that we wish to find out the population status for the species in Plantwatch and that the module includes very long-lived species the current methodology may fail to capture the smaller-scale variations which indicate response to disturbance or change, namely categories to distinguish new seedlings or regenerative growth. We can, however, use other species for the analysis to hypothesize what could have led to this peak.

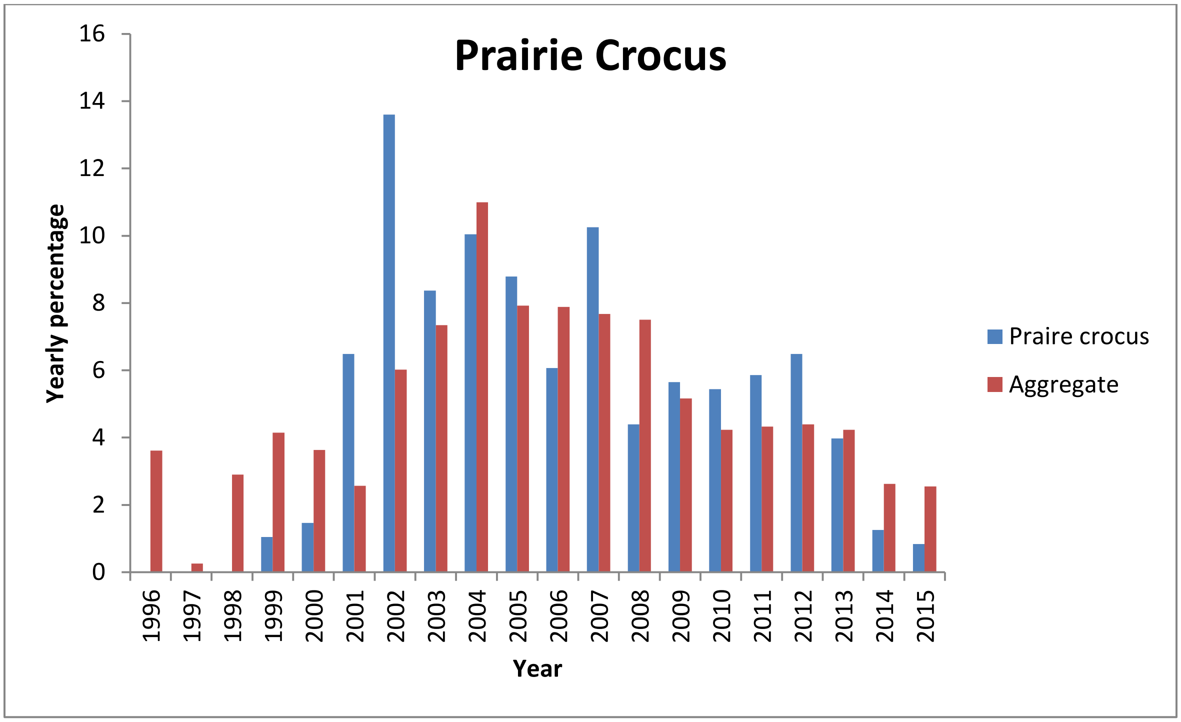

Figure 33: Prairie crocus yearly entry percentage, compared with the aggregate, from 1996 to 2015.

Figure 33 shows the data for the Prairie crocus, a species found mainly in Alberta, Saskatchewan and Manitoba – which means that its range overlaps with that of the Aspen poplar. The data indicates an enormous surge in the number of observations of this species in 2002.

Why would this simultaneous surge in the data for Aspen poplar and Prairie crocus species be relevant? This is because both species recover quickly, and in fact thrive, following forest fires (Alberta Plantwatch (2015)). This association needs to be investigated. For this we must be familiar with the time frame in which each species reacts to fire. The problem, however, is that the data in Plantwatch does not currently allow contributors to specify the development stage of observed specimens which would be vital to finding out whether the observations indicate new growth, particularly for a tree such as Aspen poplar.

In short, as it stands the Plantwatch data allows us to state that conditions in 2002 may have been conducive to Aspen poplar growth and flowering or that fires (Alberta Plantwatch, 2015, Aspen Poplar) in prior years may have led to a sudden population boom in 2002, based on our observation of a similar peak in 2002 for Prairie crocus. It does not however allow us to determine whether this is the result of new growth following disturbance, simple successful seeding or if it’s an artifact of observer bias. More research would be needed to find this information, and improving the database may be necessary to allow for the collection of this type of data in the future.

Chapter 5. Icewatch analysis

The way to plot the Icewatch data used here differs from what was used for Frogwatch and Plantwatch. As there are relatively few variables to consider, it would make sense to focus on the values over time at each location (as opposed to dividing the data by species). However this would entail producing a graph and running calculations for every location, of which there are many hundreds. This is obviously beyond the scope of this metadata analysis and would be of limited use for trend analysis. What we can do is instead focus on a more refined temporal analysis of aggregated locations instead of every single location. To do this we will find the Ordinal date for each “ice off” and “ice-on” timing This can be calculated from the visit_date column by converting it to a Julian date and subtracting the value corresponding to the year from the total.

The Julian date format used in MS Excel is a seven-digit number where the first four are the year (e.g. 2016) and the last three are the Ordinal date. The Ordinal date itself corresponds to the day in the year (from 1 to 365 or 366 for leap years), and this is used for the graphs because it is easier to plot on a graph from year to year. The formula used in excel is the following:

=B2-DATE(YEAR(B2),1,0)

B2 in this formula is the cell containing the visit date. The formula essentially involves taking the Julian date of B2 and subtracting the Julian date of “day zero” of the same year from it. In essence if B2 is April 23rd 1900, the Julian date is 1900114. The “day zero” of 1900’s Julian date is 1900000. As a result the formula essentially amounts to 1900114-1900000=114. 114 is the ordinal date of the visit. Note that the column where this formula is must be locked to “**text”** format to display the ordinal date properly.

This allows us to establish when ice forms and melts in a given year, and plot this data in graphs. By plotting the Icewatch data in a program such as ArcMap we can also focus the analysis on specific regions to see if the trends observed in the general database are visible there by selecting the locations on the map and running the analysis with these points alone.

What has been observed in general, and this will be elaborated upon further in the following section, is that IceWatch produces data that is quite useful for monitoring ice cover trends over time. While larger samples are always desirable, and participation is a requirement, unlike with FrogWatch and PlantWatch the data for this module is quite well insulated from the errors which could result from a species being cryptic or a local promotion focussing on particular species. In other words, the data is much easier to quantify as it lacks the qualitative "species" parameter which can complicate the interpretation of the data from the other 2 modules.

As such, it can be said that Icewatch shows particular promise for research on climate change.

The complete table analysis for this module is presented in Appendix C.

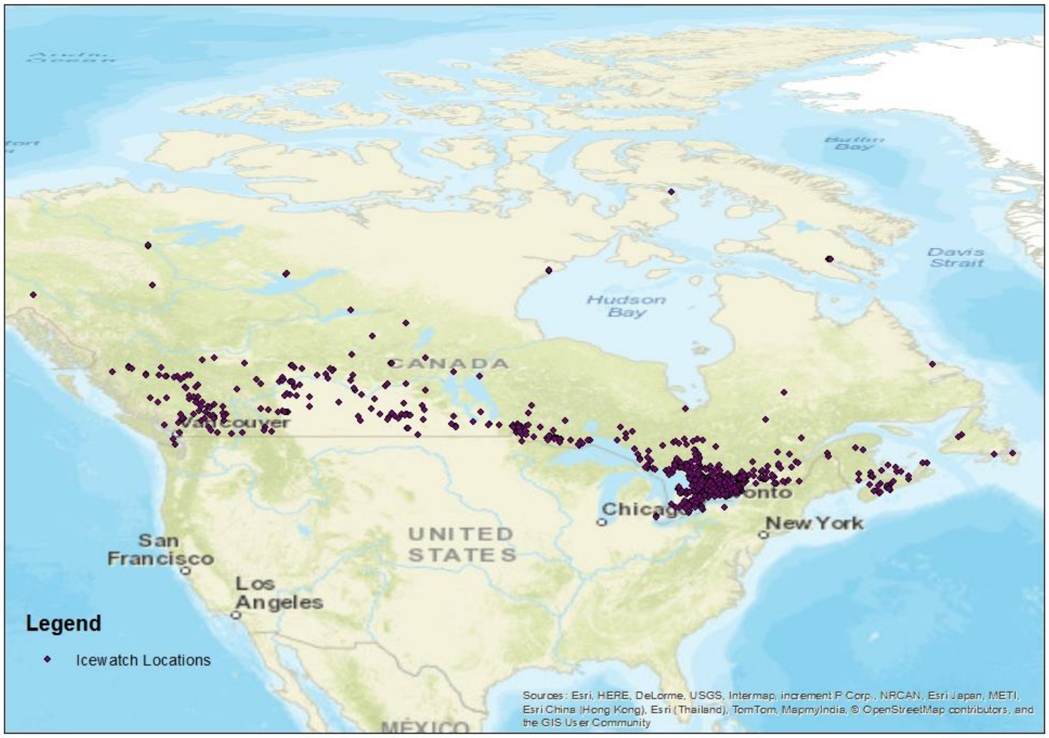

Figure 34: Map of Icewatch observation locations throughout Canada.

Figure 34 simply shows where the Icewatch observations occurred. As we can see a high proportion of the locations are concentrated around Southern Ontario (Toronto and Ottawa areas). Note that these do not each represent single observations – repeated observations at a given location do not appear differently from single instances on this map.

General trend

The relevant factors for IceWatch being much simpler than those of Frogwatch and Icewatch, especially as all data points can be compared easily (there are no "apples and oranges" situations), we do not need to divide the data further than by year in order to determine general trends. As such, by calculating the yearly mean of ice cover changes timing for every year (in this case starting in 1900) we ought to be able to produce legible graphs which will orient further analysis easily.

The data for the following graphs has been subject to some basic cleaning in order to make it easier to interpret and to eliminate outliers. In essence once the rough time frame of ice melt and formation were established those data points that fell far outside of these ranges were removed – this includes instances of ice-on reported in June, which cannot be considered for analysis of seasonal trends (assuming it actually occurred such an event is no reflection of seasonal variations, and so was removed). This also includes ice-on dates which occur very early in the year (January and February, for instance) – while these could reflect later ice formation they were removed because they would have skewed the results – and also because they may reflect refreezing after a short-term thaw. Ice-off dates which similarly occurred very late in the year may reflect fluctuating weather but would skew results.

Before we proceed, it may be appropriate to discuss the ice cover trends of certain other regions of the world to see if there is a match or at least a similar trend. In their analysis of 11 lakes in Switzerland, Franssen and Scherrer (2008) found that overall ice cover had declined over the period from 1914 to 2006. What is particularly notable is how this trend was most distinct in the last 20 years of sampling (Franssen and Scherrer, 2008). The data from this study may be of particular use to us as it was collected from a variety of sources, including newspapers, fishermen and other volunteers, and the researchers had to deal with similar issues as we may face with Naturewatch (Franssen and Scherrer, 2008).

In short, when both the ice-on and ice-off data is considered, we ought to expect a shift in ice formation or melt timings which lead to an overall shorter ice duration over the course of the 20th century.

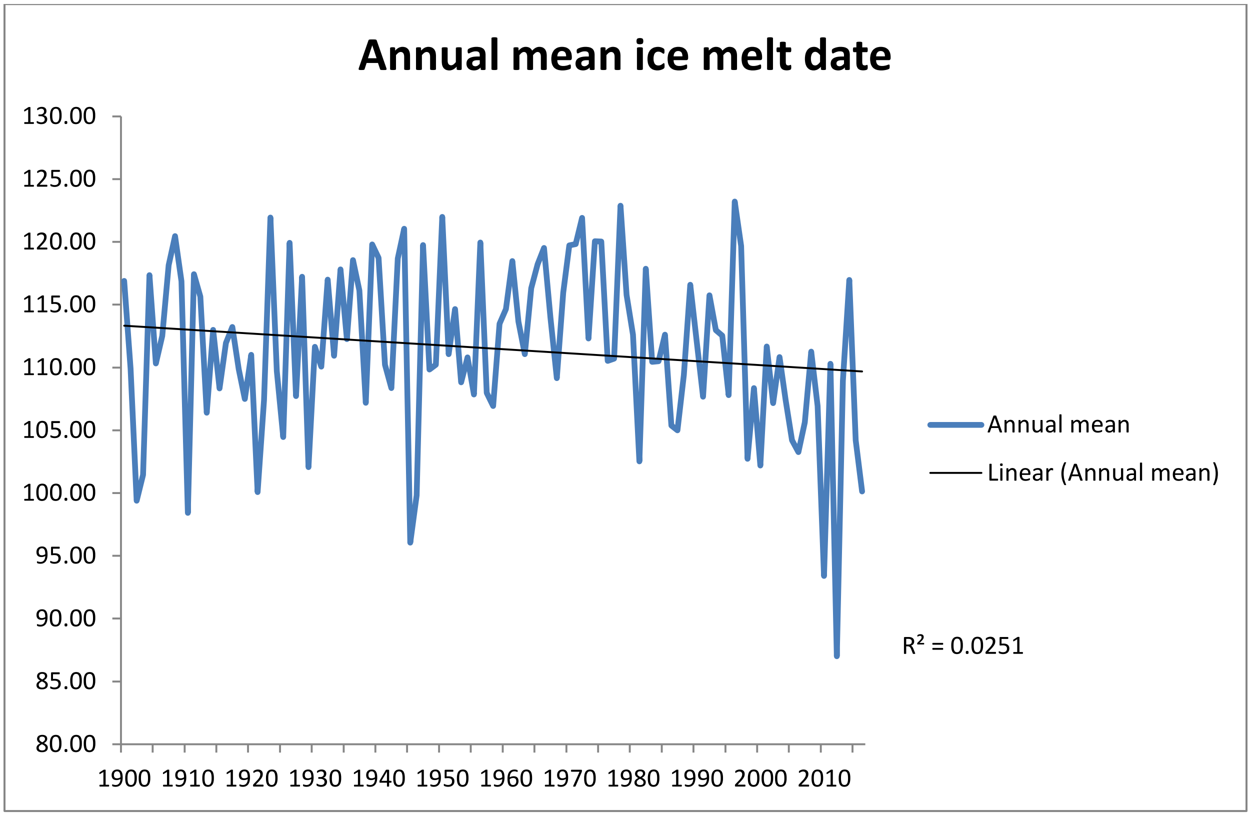

Figure 35: Ice-off average ordinal dates for every year of Icewatch.

Figure 35 illustrates that there is little change in the overall ice melt date over the course of the 20th century for the entirety of Icewatch. While it does vary from year-to-year the data on average shows very little change. There are 6083 entries in this sample.

This shows a slight trend of ice melt occurring around 4 days earlier over the course of the last century. It must be noted that this trend is hardly noticeable except perhaps in the last decade, and will not be considered relevant in this analysis.

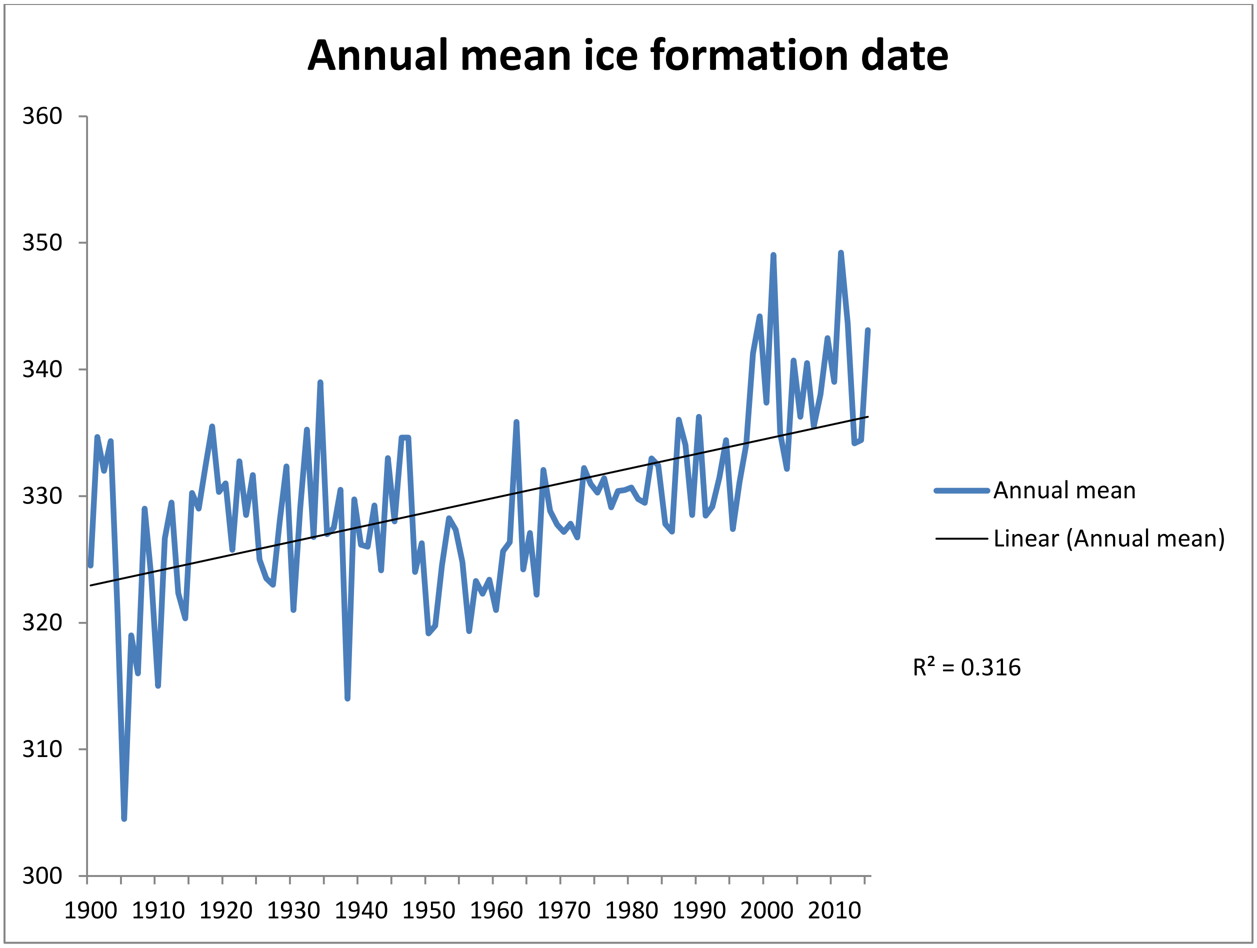

Figure 36: Ice-on average ordinal dates for every year of Icewatch.

Figure 36 shows that there is a noticeable trend for ice to form later in the year. It correlates strongly with the other evidence for global warming, however another question to answer here would be why ice formation seems to respond to the climate signal strongly while the ice thaw seems insulated from this change. There are 3332 entries for the ice-on data in this sample.

What we can see is that this could mean that the data for Icewatch correlates well with data such as that from the study on Swiss lakes (Franssen and Scherrer, 2008).

The lower number of "ice-on" entries relative to "ice-off" entries in this dataset is a pattern repeated in the subdivisions which will be analysed later. In many areas this means that the Ice-off data is much more extensive and consistent than for the ice-on. This does highlight the need to promote a greater focus on ice-on dates as these have a much greater variability and seem to be better suited to following general trends. The exact reason why ice-on data is less well covered than ice-off is not immediately clear, but it could simply be due to monitoring priorities for the older sites as the relative proportion of ice-on entries increases over time. The source of the oldest IceWatch entries especially those from the sites with the longest continuous monitoring periods is a worthwhile investigation to determine what the focus of these was. This is especially important to know as the 7 areas with the most entries (starting in 1900 for the most part, some earlier) all show no updates since the early 1990s – why they no longer contribute entries may be an important question to answer.

5.1 Icewatch newsletter

Figure 37: Icewatch module of Naturewatch Canada.

A particularly interesting question concerning the Icewatch module (Figure 37) is how useful it may be at smaller geographical scales. We shall analyse in this newsletter the Ordinal dates of ice formation and melt for Ontario is used for comparison to the general trend for the module, then follow it up with the local scale. I anticipate that the regional scale will yield the clearest results.

<u>Ontario trends</u>

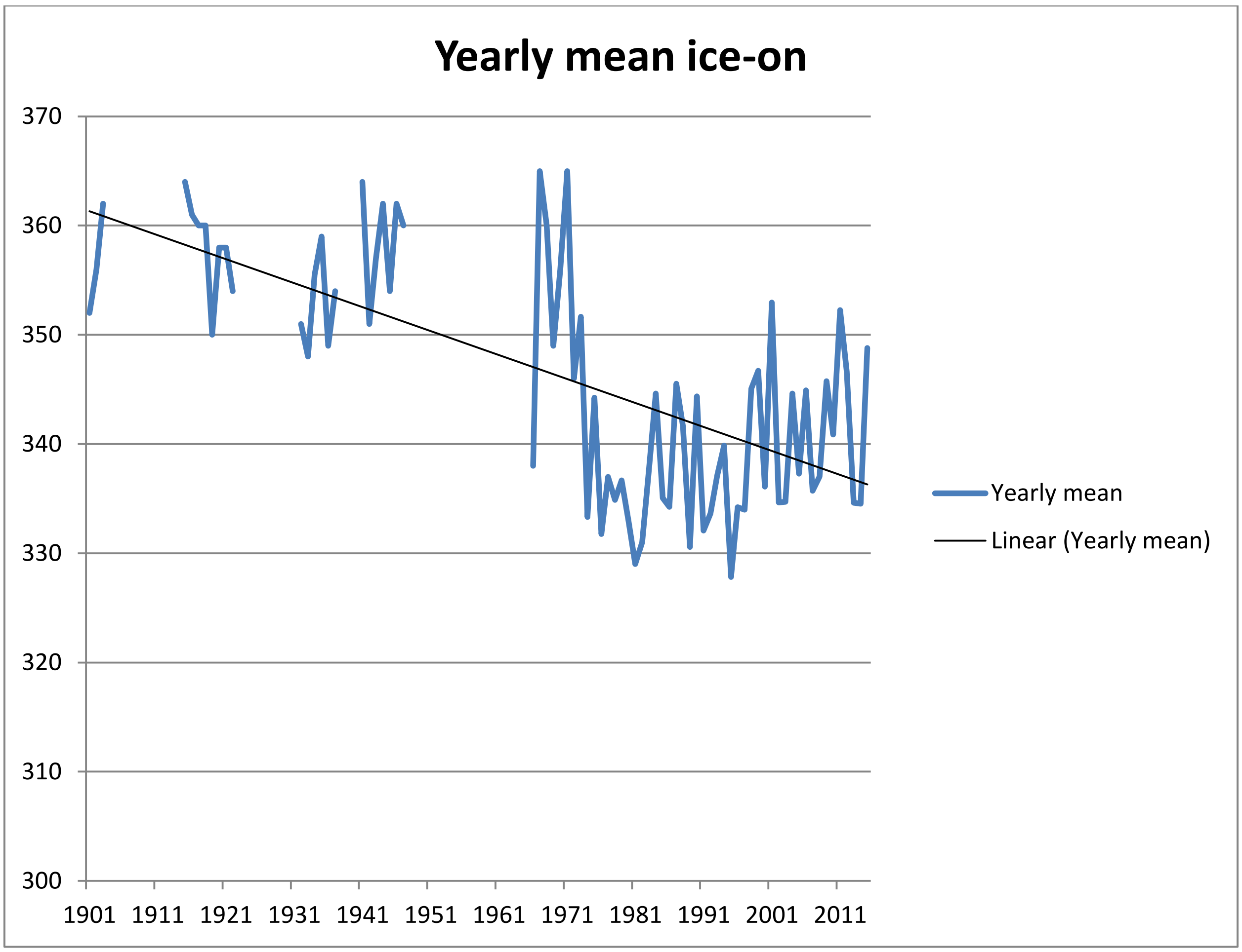

Figure 38: Ice-on average ordinal dates for available years for Ontario in Icewatch.

As we can see from Figure 38, many years are missing from this dataset. We must then limit our analysis to the continuous period for which we have entries. Trying to use the data from 1901 to 1960 roughly is simply not valid.

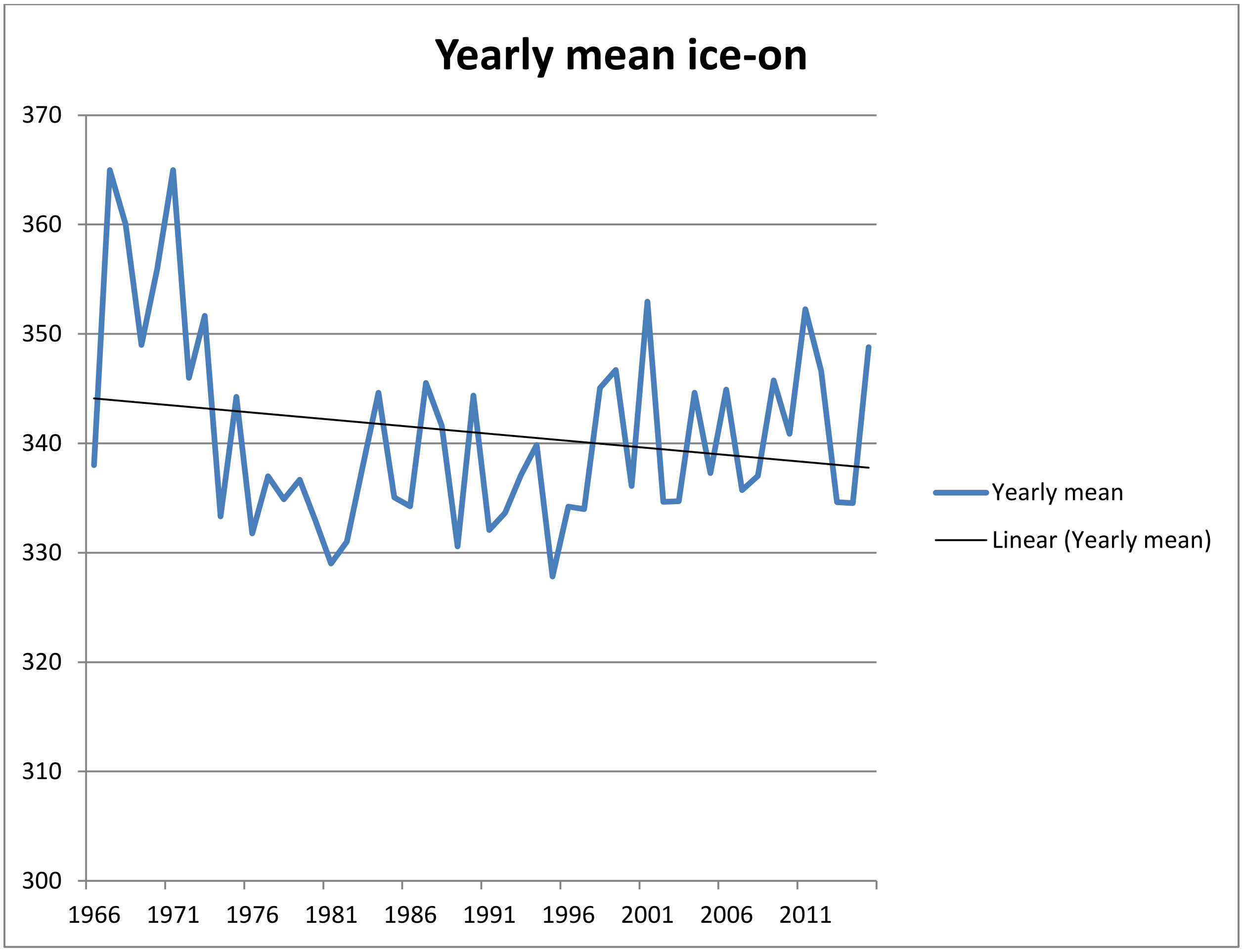

Figure 39: Ontario ice-on average yearly ordinal date, only years with continuous data.

As the Figure 39 shows, the data for ice-on is much more scattered and has poorer coverage than for the ice-off. Numbers-wise the ice-off data far outnumbered the ice-on data earlier in the dataset (this disparity is much reduced in the last 2 decades), and for the province of Ontario this leads to a gap in the chronology. A better analysis for ice-on in this region would be to only analyse the trends starting in the late 60s, which is the date from which there is data every year. There are 1691 entries in this dataset.

If we compare Figure 39 to the aggregate ice-on trend displayed in Figure 36, we can see that the trends from the national dataset are not reflected in the Ontario dataset – however this could be the result of the latitude of new sampled locations varying over time. As mentioned previously further narrowing the data to smaller regions or even single locations is where the data is least likely to be influenced by variations in latitude.

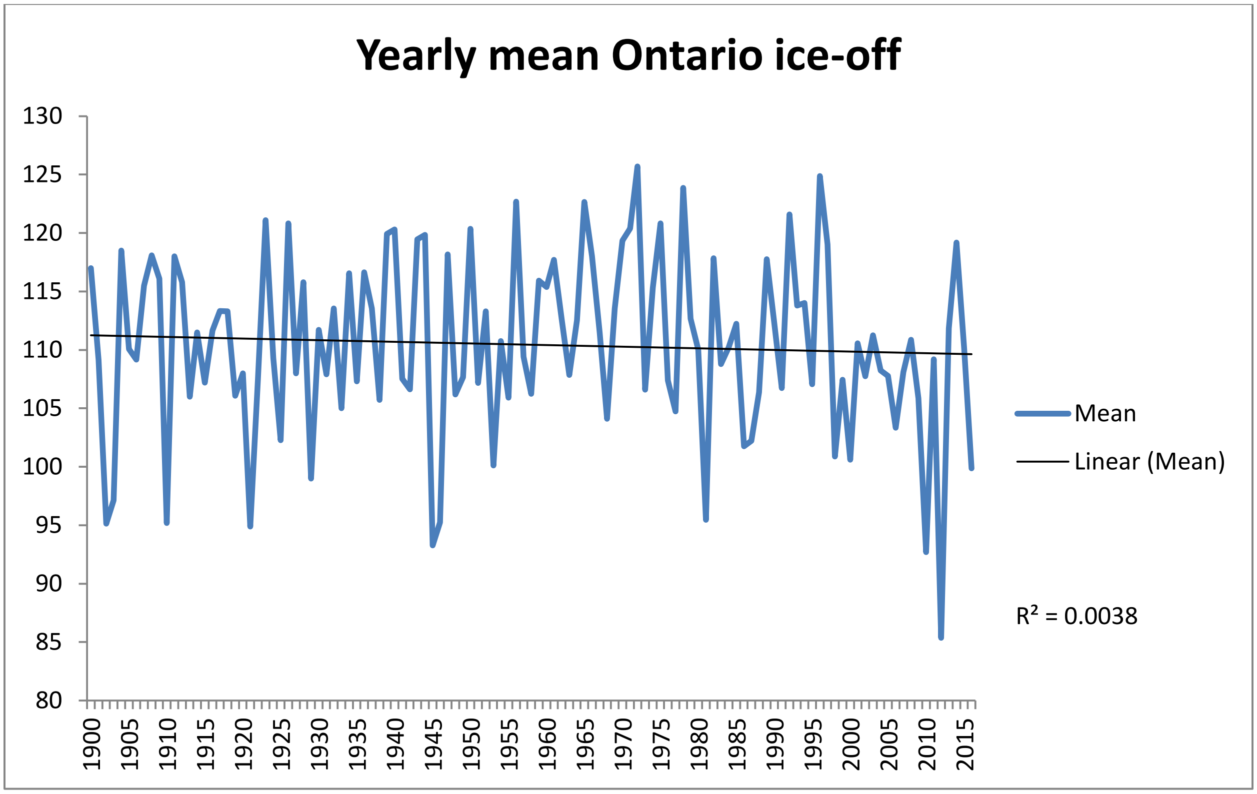

Figure 40: Ontario ice-off yearly ordinal date.

Figure 40 shows a trend remarkably similar to that of the aggregate – being a relatively constant date of ice melt. This would definitely indicate that the time frames of ice formation and melt are different, meaning that we must consider why – some water column dynamics may impact the rate at which ice forms or disintegrates over water bodies. There are 3821 entries for this dataset.

Expanding on why ice melt seems to be insulated from general climate trends to an extent may be a very worthwhile line of inquiry.

<u>Regional trends, southwestern Ontario</u>

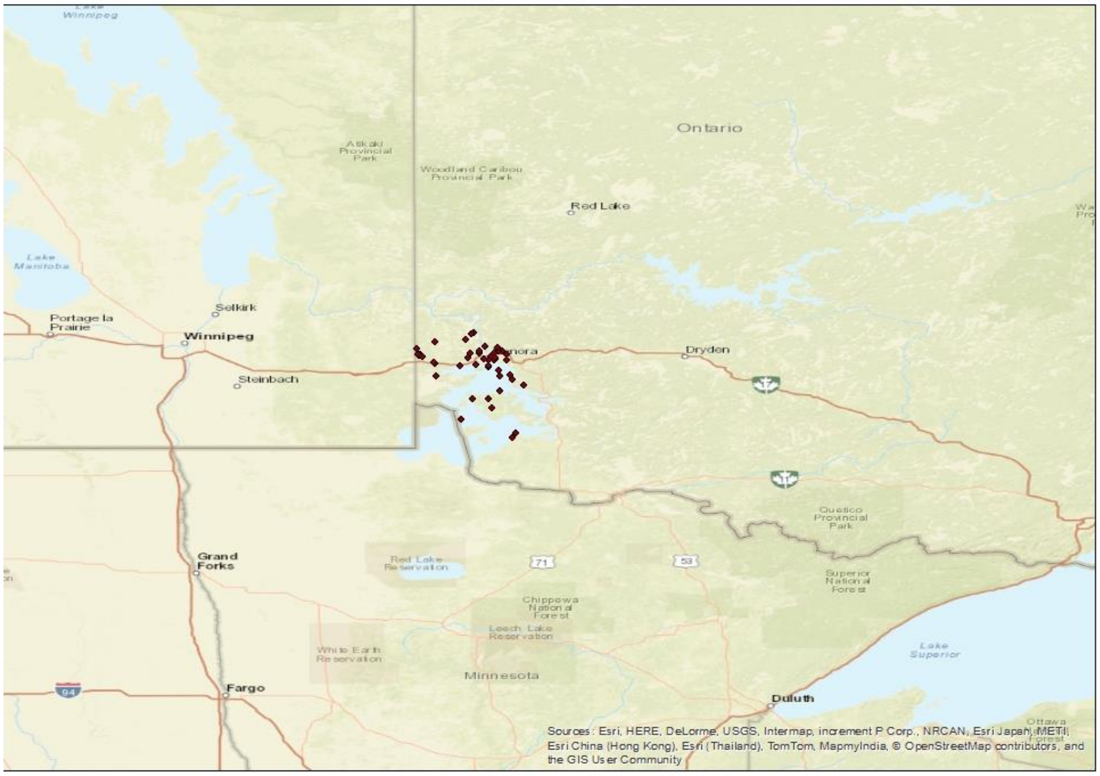

Figure 41: Map of the locations from the selected regions.

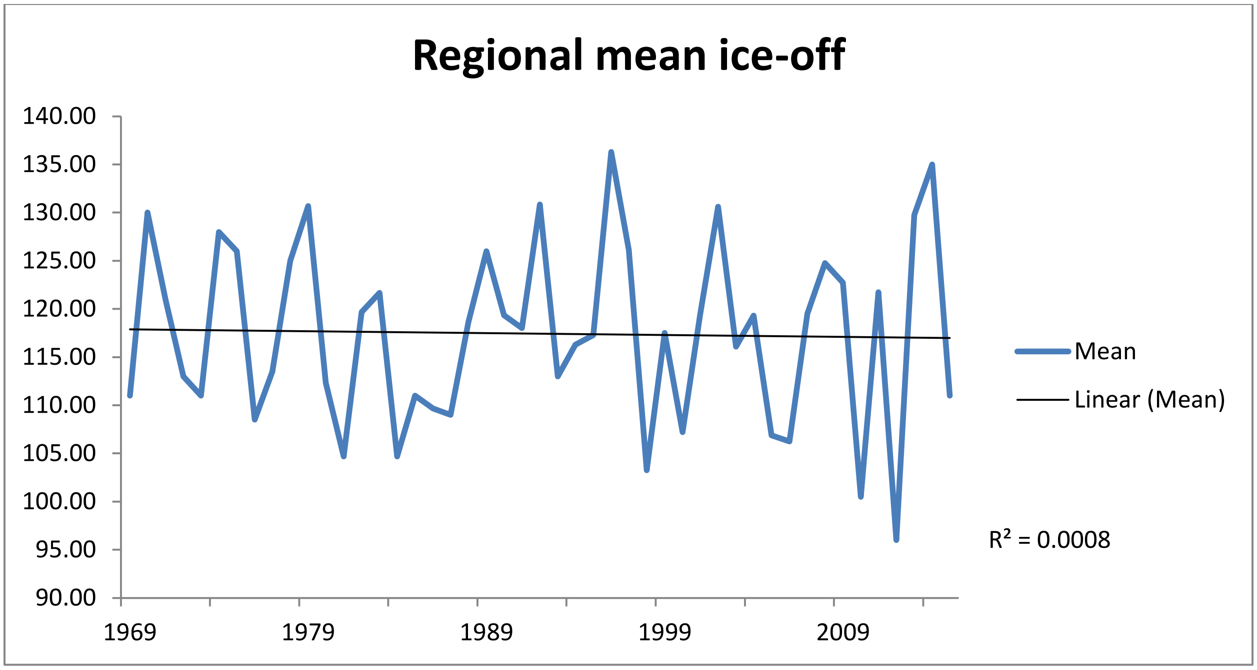

Figure 42: Yearly mean ordinal date for time of ice melt for the selected region.

Figure 42 shows a similar consistency as for the province and the aggregate. Unfortunately the data is restricted to 1969-onwards, but it may still be of use. It must be noted that this region had a total of 256 entries over this period.

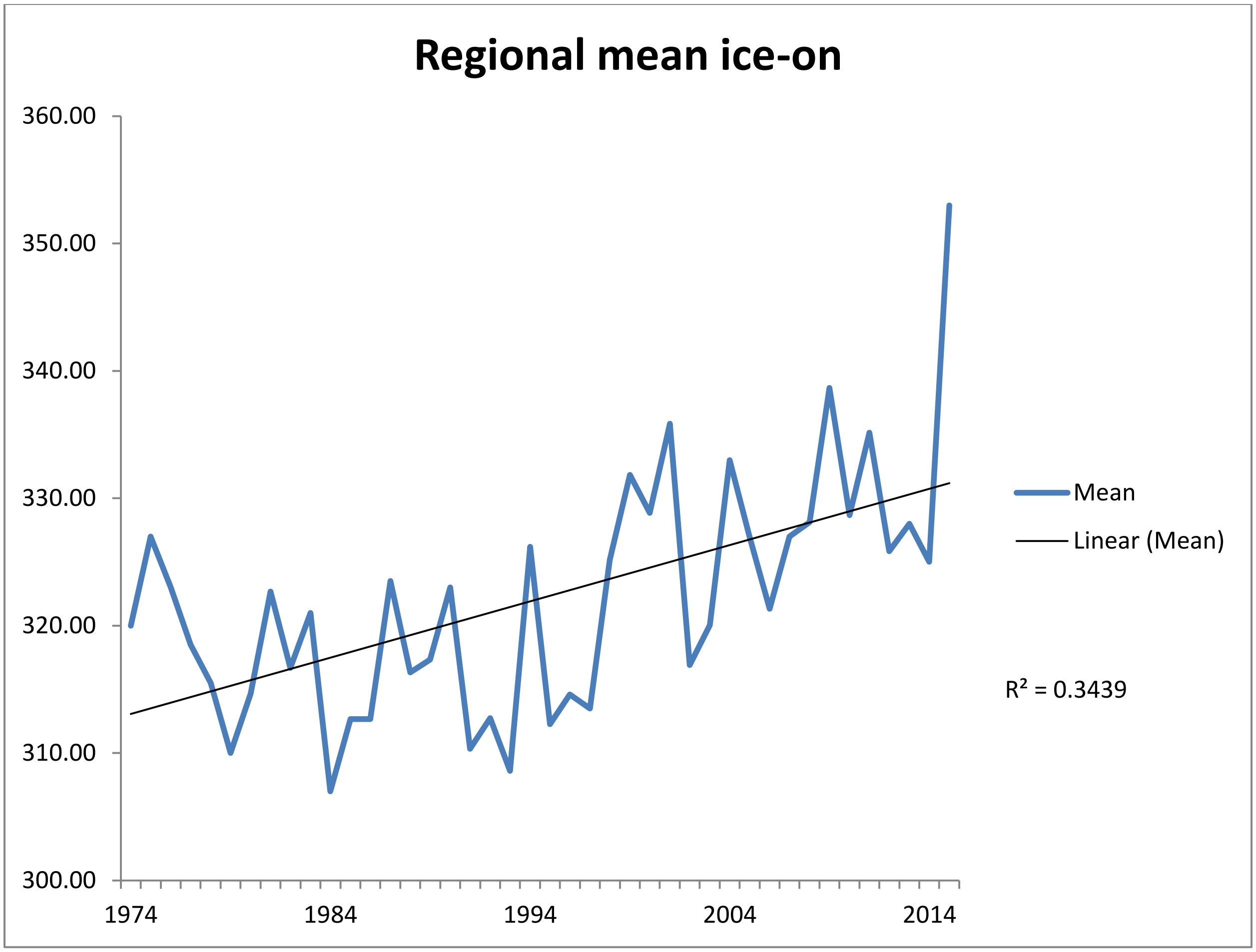

Figure 43: Ice-on yearly average ordinal date for the selected region.

As Figure 43 shows, there has been a definite trend of later ice formation in the region from 1974 to the present, with ice forming some 17 days later from 1974 to 2015.

There are 199 ice-on entries in the region. Another factor this reveals is that, in this smaller but more recent dataset, the gap between ice-on and ice-off data is not nearly as great as it is for the older provincial and aggregate data. This would support the hypothesis that the gap between ice-on and ice-off is mostly due to the monitoring practices or goals of the older monitoring initiatives. In fact, given that there is not yet any ice-on data for 2016 (while there is ice-off data), we ought to expect this gap to shrink even more as the year 2016 comes to a close, as this is the time when is ice forms over water bodies. It must be noted, however, that the most recent years for IceWatch have seen a decrease in entry counts, possibly as a result of decreased promotion of NatureWatch in general and IceWatch in particular.

Nonetheless, if the data here is significant, it indicates a delay of almost 20 days in ice formation over the course of four decades, which is most definitely a worrying trend. It would also correspond well with the aggregate data's showing the most marked shift in the later decades of the 20^{th} century.

<u>Overall data quality and value</u>

What the previous data showed is that IceWatch as a module of NatureWatch can be used to find trends in weather by tying these to its chosen proxy, being ice cover. The simplicity of the data it collects leads to data that is easy to gather and analyse as well. As a result this is an extremely valuable module for study, however its subject matter is somewhat less charismatic than FrogWatch or Plantwatch, which results in it requiring more promotion and incentives to gain and retain participants. The usefulness of the data could function as a "selling point", as the knowledge that the data will be used is one of the great motivators for citizen science.

We have learned from the Icewatch module that a focus on particular locations and how the data changes over time may provide the best results for tracking changes over time. This is especially true for Frogwatch, where the observation point is fixed in position and where successive observations can be registered by the same observer (or a group of observers). As such it may be wise for further study to focus either on a given region or to gather data from a group of known and active locations and do more refined temporal analyses for these using the ordinal dates like we have done in Icewatch.

Chapter 6. Summary

6.1 General conclusions

This metadata analysis has allowed us to discover that any study wishing to include NatureWatch as a source of data must consider what module it wishes to draw from, as the different modules do not all share the same type of utility.

We can however assert in general that using the data from Naturewatch requires some general precautions regardless of the module as well. For the most part, it is difficult to isolate entry density in any module from the effect of participation, meaning that overall without better tools or new data options population density for the species-specific modules is not quite feasible to ascertain.

What can be done as of yet is a simpler presence and absence analysis and analysing how these may shift over time and space. This means that the amount of detail in the data and its collection method does not allow for precise monitoring of local biomes and their populations. This is not surprising given what we have learned about citizen science over the course of the literature review.

The national, undivided scale of the data in itself causes problems for in-depth analyses – which means that any attempt to assess in-depth data must be done by isolating a region with a suitably high density of data points. By and large this means that Frogwatch and Plantwatch cannot be used for local analyses outside of the Toronto, Ottawa and the Great Lakes regions. Icewatch is much better for this type of analysis overall as many of its locations have several decades of data.

Combining multiple analyses can, however, allow users of the data to infer general trends. This more superficial level of information has great value for decision-makers, however, as more detailed information is generally handled by those to whom specific tasks are delegated to.

6.2 Frogwatch

The data for Frogwatch has a rather substantial sample size for certain species, but almost no entries for others. This is also the youngest module with regards to how far back the data goes.

The immense disparity in observation rates reflects two main characteristics of the module: observation heavily depends on how easy it is to find specimens (visually and audibly), and how numerous the species is. Several of the frog species, notably the Spring Peeper, are very easy to identify from its calls even though it is hard to see (being small), while others have calls which are difficult to identify or even hear at all. This leads to the overrepresentation of "notorious" species (in this case meaning easily found) in the data, which is a problem as it provides only a partial picture of the local diversity and changes in the population (or even presence) of a stealthy species can easily go entirely unnoticed. This is in part due to the nature of the monitoring done for Frogwatch – for safety reasons observers are advised not to leave familiar areas as most monitoring is done at night for frog calls.

Another reason why it may have been difficult to parse the results for trends over time is the national scale of the analyses carried out for this module. Since Frogwatch monitors elements of the biosphere for which climate is not the only critical variable, and many of the relevant variables (which may have to do with forest cover, wetland condition, human proximity, and rival species, not to mention participation) are very dissimilar throughout Canada, it stands to reason that an analysis carried out at the national level may fail to catch a clear common trend. It would make more sense to analyse individual regions or to compare regions with similar conditions – but determining if selected regions shared as many of these properties as possible was beyond the scope of this metadata analysis.

NatureWatch Canada is already a partner with several other programs for its promotion, but reaching out to other, more specialised programs would be good for potential joint initiatives and refining its monitoring capacity. A perfect example of this the Frogwatch USA program, which uses its FieldScope program to allow people to generate maps of observations, and includes tools to graph them as well. When we consider how much overlap there is between the purpose of these programs and the fact that both of these suffer from the national borders preventing some important geographical analysis, it seems that some level of cooperation ought to take place. What forms this cooperation takes may be to found a separate joint database with both of the combined data and whatever functionalities that can be supported, or it could entail sharing knowledge of the methodologies and software used. The second option would be aimed at improving the utility of both (but likely Naturewatch's Frogwatch would benefit the most from such an exchange). It would also make further cooperation easier as it would bring the two databases closer, allowing for the formation of a higher-quality joint database than would have previously been possible.

6.3 Plantwatch

Most of the points for Frogwatch apply for Plantwatch, except of course for the fact that most observations in plantwatch are visual rather than audible. As such, the same issues arise with those species which are easier to find due to visibility or proximity to human population centers being over-represented in the data.

The observation method differs, however, in that Plantwatch requires observers to move through the location to find specimens as opposed to establishing a "stakeout" and listening to calls. As a result consecutive observations can be subject to changes due to new paths being taken, unless the contributor consciously remains on already-monitored areas.

A second and considerable issue with the Plantwatch modules is that it is currently not particularly well equipped to monitor species with longer life cycles, or trees in particular, as we have discussed in the Plantwatch newsletter.

6.4 Icewatch

In many regards, Icewatch is the simplest of the datasets – it only has 2 important factors, being ice status and date of change. It is also the module with the oldest data – some of its entries go back to the start of the 20th century, and a few even before then. The relation between the entries and their location is much tighter for this module than the others, and it also depends more strongly on repeated observations from year to year. This means that the data for icewatch is generally well-suited for finding temporal trends – indeed it is quite simply the most convenient for this at the local, provincial and national scale.

To note, the most useful data for tracking changes over time are those for ice formation rather than ice melt. This is an issue for some of the older data as the ratio of ice-off entries to ice-on entries was very high early on even though it has been decreasing over time. Promoting Icewatch should include mentions that ice formation dates are very important to the analysis to make sure this apparently more sensitive variable is better monitored.

Besides the general properties of the data itself, the brief analysis carried out (detailed in the Icewatch newsletter) shows that ice-on data shows a strong trend for ice to form later in the year over the course of the 20th century.

6.5 Recommendations for future work

Given what we know of the pros and cons of this database, along with the trends it has so far allowed us to see, it would make sense to consider what could be done to improve it.

Naturally the first recommendation is to further promote the database. As with any research sample, having a larger sample allows for more relevant results and for filtering out outliers. This is most important for Frogwatch and Plantwatch, as these modules require much more extensive data to be used.

As we have discussed, Icewatch's focus on continued monitoring of a particular location/water body makes monitoring trends over time much easier and relevant. Placing greater emphasis on a similar methodology for Frogwatch and Plantwatch may allow for future research to be focussed on regions rather than species at the national scale– which may be the key to finding clearer trends in those modules.

Another thing which may be of interest to researchers is to use the ordinal date for observations in the Frogwatch and Plantwatch modules, and so track the dates of observations for a multitude of variables. This could be, for instance, tracking when the Gray's Treefrog observations in Ontario occurred over the years, and seeing if the calls were heard earlier or later in the year for this species in this province. This methodology could be used as a means of refining the understanding of the times where a given species may call, which would become a new analysis baseline. In other words, new observations coming in could be sorted by location and ordinal date first, and if a clear "baseline" has been set in previous years (which species, and when roughly) this could be a way to spot new developments (species showing up or not, or being observed/heard outside of the normal period).

Given how important the ordinal date is for plotting the Icewatch data, and how it may prove similarly valuable for Frogwatch and Plantwatch, it may be a good idea to have a column that calculates this value automatically for all entries, including the new ones – this would facilitate the process since some datasheet programs may not allow this operation to be done as easily as it is in Excel.

=B2-DATE(YEAR(B2),1,0)

When it comes to geographical distribution, the data suffered most from one of the key shortcomings we predicted following the literature review: concentration in the southern latitudes of Canada, and more specifically in and around densely populated areas. This shortcoming would require greater monitoring at the more remote areas of the North, although the number of potential observers in these areas who could contribute to Naturewatch is rather low both because there are few people but also because of limited infrastructure to support things like Internet, which precludes the individual observers from doing so. Finding the coordinates of sampled locations may be more difficult there as well.

A different way to address the issues of verifiability for citizen science, most pertinent for Icewatch, would be to combine analyses of Icewatch observations with remote sensing studies focussed on the same water bodies and the timing of freezing and thaw. The field of remote sensing is gaining increasing traction and it is another means of acquiring data over wide areas for fewer resources.

One possibility for addressing this problem would be to develop a methodology or set of tools which would allow communities or groups in these areas to gather their observations in paper format where required and compile them to be later put into the database. This would not only provide a means of making up for the current lack of entries at these latitudes but it may also be a way to involve different communities, including aboriginal ones. Since many communities in the North struggle to have their voices heard, for a multitude of reasons, this may contribute to bettering their living conditions – and contributing to a database which is being used for research may be a first step to greater involvement.

References

Beaubien, E.G., Hamann, A. (2011 a) "Plant phenology networks of citizen scientists: recommendations from two decades of experience in Canada" *International Journal of Biometeorology* (55) p. 833-841

Beaubien, E.G. and Hamann, A. (2011 b) "Spring Flowering Response to Climate Change between 1936 and 2006 in Alberta, Canada". *BioScience* (61, No. 7) p. 450-459

Bliss, John., Aplet, Greg., Hartzell,Cate., Harwood, Peggy., Jahnige, Paul., Kittredge, David., Lewandowski, Stephan., Soscia, Mary Lou., (2001) "Community-Based Ecosystem Monitoring" *Journal of Sustainable Forestry* (12, #3-4) p.143-167

Cathy C. Conrad, Cathy C., Hilchey, Krista G. (2011) "A review of citizen science and community-based
environmental monitoring: issues and opportunities" Environmental Monitoring and Assessment (176) p. 273-291

Chavis, David M., Stucky, Paul E., Wandersman, Abraham., (1983) "Returning Basic Research to the Community: A Relationship Between Scientist and Citizen" *American Psychologist* (38 #4) p.424-434

Crain, Rhiannon., Cooper, Caren., Dickinson, Janis L. (2014) "Citizen Science: A Tool for Integrating Studies of Human and Natural Systems" *Annual Review of Environment and Resources* 39 p. 641-645

Crowston, Kevin., Prestopnik, Nathan R. (2013) "Motivation and data quality in a citizen science game: A design science evaluation" *2013 46th Hawaii International Conference on System Sciences* p.450-459

Devictor, Vincent., Whittaker, Robert J., Beltrame, Coralie. "Beyond scarcity: citizen science programmes as useful tools for conservation biogeography" *Diversity and Distributions* (2010) 16, 354–362

Dunn, E. H., Francis, C. M., Blancher, P. J., Drennan, S. R., Howe, M. A., LePage, D., et al. (2005).
Enhancing the scientific value of the Christmas Bird Count. Auk, 122, 338–346. doi:10.1642/0004-
8038(2005)122[0338:ETSVOT]2.0.CO;2.

Edwards, Richard. (2014) "The 'Citizens' in Citizen Science Projects: Educational and Conceptual Issues" *International Journal of Science Education* (4, #4) p. 376-391

Foody, Giles M. (2014) "Rating crowdsourced annotations: evaluating contributions of variable quality and completeness" *International Journal of Digital Earth* (7, #8) p. 650-670

Franssen, H. J. Hendricks., Scherrer, S. C. (2008) "Freezing of lakes on the Swiss plateau in the period 1901–2006" *International Journal of Climatology* (28) p. 421-433

Grimm, E. (2008) "Neotoma: An Ecosystem Database for the Pliocene, Pleistocene, and Holocene" *Illinois State Museum Scientific Papers E Series 1* p. 1-94 http://www.neotomadb.org/uploads/NeotomaManual.pdf

Henden, Arne A. (2011) "Amateur Community and "Citizen Science" *New Horizons in Time-Domain Astronomy Proceedings IAU Symposium No. 285, 2011 R.E.M. Griffin, R.J. Hanisch & R. Seaman, eds.* p.255-260

Hiller, Michelle A. (1991) "The National Estuary Program and Public Involvement" *Marine Pollution Bulletin* (23) p. 645-648

Hunter, Jane., Alabri, Abdulmonem., van Ingen, Catharine. (2013) "Assessing the quality and trustworthiness of citizen science data" *Concurrency Computation: Practice and Experience* (25) p.454–466

Keen, A. (2007). The cult of the amateur: How today's internet is killing our culture and assaulting our economy. London: Nicholas Brealey.

Mansell, Robin. (2014) "Empowerment and/or Disempowerment: The Politics of Digital Media" *Popular Communication* (12) p. 223-236

Mansuri, Ghazala., Rao, Vijayendra., (2004) "Community-Based and -Driven Development: A Critical Review" *The World Bank Research Observer* (19, #1) p.1-39

Menzel A, Estrella N, Testka A. (2005) "Temperature response rates from longterm phenological records" *Climate Research* (30) p.21–28.

Menzel A, et al. (2006) "European phenological response to climate change matches the warming pattern" *Global Change Biology* (12) p.1969–1976.

Payne, Christine A., (1999) "The Challenges of Employing performance monitoring in public health community-based efforts: a case study" *Journal of Community Health (* 24, #2) p.159-169

Purdam, Kingsley., (2014) "Citizen social science and citizen data? Methodological and ethical challenges for social research" *Current Sociology* (62 #3) p.374–392

Ranard, Benjamin L., Ha, Yoonhee P., Meisel, Zachary F., Asch, David A., Hill, Shawndra S., Becker, Lance B., Seymour, Anne K., Merchant, Raina M., (2013) "Crowdsourcing—Harnessing the Masses to Advance Health and Medicine, a Systematic Review" *Journal of General Internal Medicine (*29 #1) p.187-203

Rosenzweig C, Casassa G, Karoly DJ, Imeson A, Liu C, Menzel A, Rawlins S, Root TL, Seguin B, Tryjanowski P. (2007). "Assessment of observed changes and responses in natural and managed systems" Pages 79–131 in Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hanson CE, eds. Climate Change 2007: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Simula, Henry., (2013) "The Rise and Fall of Crowdsourcing?" *46th Hawaii International Conference on System Sciences* p.2783-2791

Stevens, Matthias., Vitos, Michalis., Altenbuchner, Julia., Conquest, Gillian., Lewis, Jerome., Haklay, Muki., (2014) "Taking Participatory Citizen Science to Extremes"

Stewart, O., Lubensky, D., & Huerta, J. M. (2010). Crowdsourcing participation inequality: A SCOUT model for the enterprise domain. in Proceedings of the ACM SIGKDD Workshop on Human Computation, pp. 30–33.

Thornton, Teresa., Leahy, Jessica., (2012) "Trust in Citizen Science Research: a case study of the groundwater education through water evaluation & testing program" *Journal of the American Water Resources Association* (48, #5) p.1032-1040

Tippins, Deborah J., Jensen, Lucas John., (2012) "Citizen science in digital worlds: the seduction of a temporary escape or a lifelong pursuit?" *Cultural Studies of Science Education* (7) p. 851–856

Tulloch, Ayesha I.T., Possingham, Hugh P., Joseph, Liana N., Szabo, Judit., Martin, Tara. (2013) "Realising the full potential of citizen science monitoring programs" *Biological Conservation* (165) p. 128-138

Vasseur L, Guscott RL, Mudie PJ (2001) Monitoring of spring flower phenology in Nova Scotia: comparison over the last century. Northeast Nat 8:393–402

Whitelaw, Graham., Vaughan, Hague., Craig, Brian., Atkinson, David., (2003) "Establishing the Canadian community monitoring network" *Environmental Monitoring and Assessment* (88) p.409–418

Ye, Qinghua., Wei, Qiufang., Hochschild, Volker., Duguay, Claude R. (2011) "Integrated observations of lake ice at Namco on the Tibetan Plateau from 2001 to 2009" *International Geoscience and Remote Sensing Society* (IGARSS), p.3217 - 3220

Yu, Jun., Wong, Weng-Keen., Hutchison, Rebecca A. (2010) "Modeling Experts and Novices in Citizen Science Data for Species Distribution Modeling" *2010 IEEE International Conference on Data Mining* p.1157 - 1162

Zhao, Yuxiang., Zhu, Qinghua., (2014) "Evaluation on crowdsourcing research: Current status and future direction," *Information Systems Frontiers*, (16) p. 417-434

Alberta Plantwatch (2015). Aspen Poplar. Retrieved from
http://plantwatch.naturealberta.ca/choose-your-plants/aspen-poplar/

Alberta Plantwatch (2015). Prairie Crocus. Retrieved from
http://plantwatch.naturealberta.ca/choose-your-plants/prairie-crocus/

ECCC (Environment and Climate Change Canada) (2016). The Top Ten Canadian Weather Stories for 2001. Retrieved from https://www.ec.gc.ca/meteo-weather/default.asp?lang=En&n=1DE844DD-1

National Audubon Society (2016). Christmas Bird Count. Retrieved from
http://birds.audubon.org/christmas-bird-count

National Snow and Ice Data Center (2016). Global Lake and River Ice Phenology. Retrieved from https://nsidc.org/data/lake_river_ice/freezethaw.html

Naturewatch (2016). Plant Descriptions. Retrieved from
https://www.naturewatch.ca/plantwatch/how-to-guide/plant-descriptions/

Weatherspark (2016). Historical Weather For 2001 in Toronto, Ontario, Canada. Retrieved from
https://weatherspark.com/history/28448/2001/Toronto-Ontario-Canada

Appendix A: Frogwatch analysis

Appendix A-1 Frogwatch tables

It is to be noted that the tables are listed in this section based on where they appear in the first table, being the Frogwath_observation table. The tables not referenced in the main one will appear in order of how close they are "hierarchically" to the frogwatch_observation table.

It must also be noted that reference types are not mentioned because all references in the database are made in the utf8_general_ci style. As such including this information in all tables as originally planned in the project's methodology would be redundant. The only information required is whether a particular column in the database refers to another table (and which table is referenced).

Foreigh key column headings will only be described in detail in their respective table (where they are primary keys), identified in the "reference table" column.

Several of the primary key data are also identified as being "auto-increment". This refers to how the new values are added: each different instance is assigned a value one higher than the last, simply.

1.1 Frogwatch_observation

The table where the entries for Frogwatch themselves are stored.

Table: Frogwatch_observation		
Column heading	**Data type**	**Reference table**
Pk_frogwatch_observation	Integer – auto-increment	
Fk_location	Integer	Location
Fk_observer	Integer	Observer
Fk_frog_species	Varchar(2)	Frog_species
Fk_observation_code	Varchar(10)	Frogwatch_observation_code
Create_date	Datetime	

Change_date	datetime	
Visit_date	date	
Air_temperature	Varchar(10)	
Water_temperature	Decimal(6,3)	
Comments	text	
Possible_outlier	Varchar(20)	
Validated	Varchar(20)	
Audit_date	Date	
Image	Varchar(255)	
Fk_administrator_auditor	integer	Missing
Possible_pseudacris_triseriata	Varchar(1)	

Pk_frogwatch_observation: This column assigns a number to the entries loosely based on time of entry. The auto-increment indicates that the process of assignment is automatic (increments of 1) for each new entry.
Create/change_date: Date when the observation was entered into the database, or edited.
Visit_date: Date when the observation was made, as defined by the observer.
Air/water_temperature: Where relevant, includes basic information on temperature. Mostly defunct.
Comments: Personalised comment.
Possible_outlier: Identifies possible outliers.
Validated:
Audit_date: Defunct column.
Image: Picture image file, if any.
Fk_administrator_auditor: This foreign key is not found anywhere else in the database, and all entries are NULL. Defunct table.
Possible_pseudacris_triseriata: Indicates whether the specimen COULD be Pseudacris triseriata (Western chorus frog), with a Y (yes) meaning it could possibly be and a N (No) meaning it cannot be confused with the species being observed. Mostly concerns Boreal chorus frog entries.

1.2 location

Table: location		
Column heading	**Data type**	**Reference table**
Pk_location	Integer – auto-increment	
Fk_observer	Integer	Observer
Create_date	datetime	
Change_date	datetime	
Location_name	Varchar(255)	
Observer_location_code	Varchar(255)	
Nearest_named_place	Varchar(255)	
Nnp_direction	Varchar(2)	
Nnp_distance	Integer	
latitude	Decimal(18,12)	
longitude	Decimal(18,12)	
Coordinate_source_code	Varchar(2)	
Coordinate_accuracy_code	Varchar(3)	

Location_description	Varchar(255)	
comments	Varchar(255)	
Location_type_code	Char(1)	
elevation	Varchar(20)	
province	Varchar(50)	

Pk_location: Numerical code assigned to each new location, as described in this chapter's introduction.
Create/change_date: Date at which the location was entered into the database and assigned its code.
Location_name: Name of the location as provided by the observer. This can be a general location such as the nearest city or a very specific location such as "[person]'s backyard".
Observer_location_code: Defunct column – none of the entries have non-null values. This is because observers were formerly associated with a single location in the past, while now observers are considered wholly independent from the locations where their observations are made.
Nearest_named_place: This column allows the observer to make locating where the observation was made in the event it is some distance from the "nearest named" place. Usually used in a complimentary manner with the location_name in the event it is too specific/local.
Nnp_direction: Direction to the nnp (nearest named place), using cardinal points.
Nnp_distance: Distance to the nnp.
Latitude/longitude: Coordinates provided by the observer.
Coordinate_source_code: Source of the coordinate data, as provided by the observer. It must be noted that only 191 of the locations have a value entered into this column, out of over 24 500 locations, meaning the coordinates cannot be verified with certainty.
Coordinate_accuracy_code: All values are NULL. Defunct column.
Location_description: Observers can describe the location in this column to provide some additional information.
Comments: The observer can add information here that isn't included in the other columns. All values are NULL, however.
Location_type_code: 1-digit letter code assigned to the location. Only provided in 100 locations out of the over 24 500, and the data itself is not clearly identified. Defunct column.
Elevation: Elevation at which the observation was made – mostly defunct.
Province: Name of the province (or territory) where the observation was made.

1.3 Observer

Table: observer		
Column heading	**Data type**	**Reference table**
Pk_observer	Integer – auto-increment	
Fk_partner	Integer	Partner
Fk_age_range	Integer	Age_range
Observer_code	Varchar(255)	
Observer_language	Varchar(1)	
Create_date	Datetime	
Change_date	Datetime	
Observer_first_name	Varchar(255)	
Observer_last_name	Varchar(255)	
Mailing_address_street	Varchar(255)	
Mailing_address_city	Varchar(255)	
Mailing_address_province_code	Varchar(2)	

Mailing_address_country_code	Varchar(3)	
Mailing_address_postal_code	Varchar(255)	
Telephone_number	Varchar(255)	
Email_address	Varchar(128)	
Affiliation	Varchar(128)	
Comments	Varchar(254)	
Send_newsletter	Varchar(10)	
Password	Varchar(512)	
Password_type	Varchar(50)	
Password_temporary	Varchar(16)	
Email_temp	Varchar(255)	
Temp_password_sent	Varchar(255)	
Old_observer_code	Varchar(255)	
Email_validated	Tinyint(4)	

Pk_observer: This column is a numbered identifier for each contributor/observer. The increments are 1 per observer, meaning every observer whose information does not match a pre-existing one is assigned a number 1 higher than the previous one.
Observer_code: Largely analogous to the pk_observer column on a one-for-one basis, except the starting point is much higher (1111111 as opposed to 21).
Observer_language: Possible entries are E and F, for English and French. NULL entries occur when no language is specified.
Create/change_date: Refers to the date the observer was entered into the database, or edited.
Observer_first_name, Observer_last_name: Where provided, identifies the observer's full name via these two columns.
Mailing_address_/city/province_code/country_code/postal_code: Provides the respective mailing address for each category where provided.
Telephone_number, email_address: Where provided, contains the observer's contact information.
Affilitation: The observer's affiliation to an organisation, which can be a private company, corporation or non-profit organisation, among others.
Comments: Personalised comment, where applicable.
Send_newsletter: Notes whether the observer wishes to receive newsletters published on NatureWatch.
Password, Password_type, Password_temporary: The observer's password to sign into the website, for access to the data and to enter observations. The temporary password is the one sent automatically to the observer upon registration, where the "password" variable is null unless the observer changed it to a personalised password.
Email_temp: Temporary password automatically sent to the observer upon registration.
Temp_password_sent: This column informs on whether the observer received a confirmation email for his/her registration as an observer for NatureWatch.
Old_observer_code: The database was altered over the years, including a streamlining of the process of assigning numbers to observers. This column contains the numbers for the observers prior to the update, and so more recently added observers do not have a number in this column.
Email_validated: Denotes if the observer received and acknowledged the confirmation email for the creation of his or her account on the NatureWatch website.

1.4 Frog_species

The table where the 2-digit code is equated with the species name (English, French and scientific)

Table: Frog_species		
Column heading	**Data type**	**Reference table**
Pk_frog_species	Integer	
English_name	Varchar(255)	
French_name	Varchar(255)	
Scientific_name	Varchar(255)	

Pk_frog_species: The 2-digit code given to entries based on the species observed.
English_, French_, scientific_ name: The names of the species identified by the pk_frog_species integer in the rest of the database.

1.5 frogwatch_observation_code

Table: Frog_observation_code		
Column heading	**Data type**	**Reference table**
Pk_frogwatch_observation_code	Integer	
English_name	Varchar(255)	
French_name	Varchar(255)	
English_description	Varchar(255)	
French_description	Varchar(255)	

Pk_frogwatch_observation_code: The code identifying the circumstances in which the observation was made.
English/French_ name/description: A name and description of the circumstances the code corresponds to.

1.6 Partner

Table: partner		
Column heading	**Data type**	**Reference table**
Pk_partner	Integer	
Create_date	Datetime	
Change_date	Datetime	
English_name	Varchar(255)	
French_name	Varchar(255)	
Mailing_address_street	Varchar(255)	
Mailing_address_city	Varchar(64)	
Mailing_address_province_code	Varchar(2)	
Mailing_address_country_code	Varchar(3)	
Mailing_address_postal_code	Varchar(6)	
Telephone_number	Varchar(10)	
Fax_number	Varchar(10)	

Email_address	Varchar(128)	
Contact_name	Varchar(128)	
Comments	Varchar(255)	

Pk_partner: Number code assigned to the program partner.
Create/change_date: Date when the pk_partner entry was entered or its details changed.
English/French_name: English and French names of the partner organisation.
Mailing_address_street/city/province_code/country_code/postal_code: Coordinates of the partner.
Telephone/fax_number, email_address: Contact information.
Comments: Any personalised comment, where relevant.

It must be noted that there are only 4 entries in this table, none of which contain any information besides the pk_partner and create/change date. The table as a whole is thus defunct.

1.7 age_range

Table: age_range		
Column heading	**Data type**	**Reference table**
Pk_age_range	Integer	
Range_text	Varchar(50)	

Pk_age_range: Assigns an increment code to the age of the observer.
Range_text: The values (high and low) delimiting the above increments.

1.8 habitat_type

Table: habitat_type		
Column heading	**Data type**	**Reference table**
Pk_habitat_type	Integer	
Create_date		
Change_date		
English_name		
French_name		
English_description		
French_description		

Pk_habitat_type: The number code for a particular type of habitat.
Create/Change_date: Date at which the pk_habitat_type value was entered, and that of the most recent edits to the information.
English/French_ name/description: A name and description of the habitat the code corresponds to.

Appendix A-2 Frogwatch trend analysis

Most of the trend analysis has been done through Excel. While we could have used R to pursue specific regression, in this case we have begun with year count analysis through excel in order to plot the evolution of entry numbers over the program's history. In order to facilitate comparisons between different tables, I have subsequently calculated what percentage of the total number of entries occurred each year. In this way it is possible to plot tables with vastly different entry totals (such as the aggregate and a single species), as the values are expressed as fractions of a fixed total (100%). This basic operation also required the elimination of those values without a valid date column entry (and consequently no year value).

The reason for doing the analysis with Excel was because of how each step could be tracked in the files. It also allowed us to easily see if any mistakes had been made as different data points could be isolated in the datasheets – this includes, for instance, a moment where the year was extracted from the wrong date-time column, using the posting (or entry date) rather than the visit_date which corresponds to when the observation was made.

This information will not be repeated for every graph to save space. It is also to be noted that to lighten the text specifying that the comparisons between the aggregate and a given species refer to the percentage all the time. Even if it is worded as similar entry count values, we are describing the yearly percentages, as obviously the aggregate contains many more total entries. This was to avoid repeating the same sentences over and over again when an analysis is carried out.

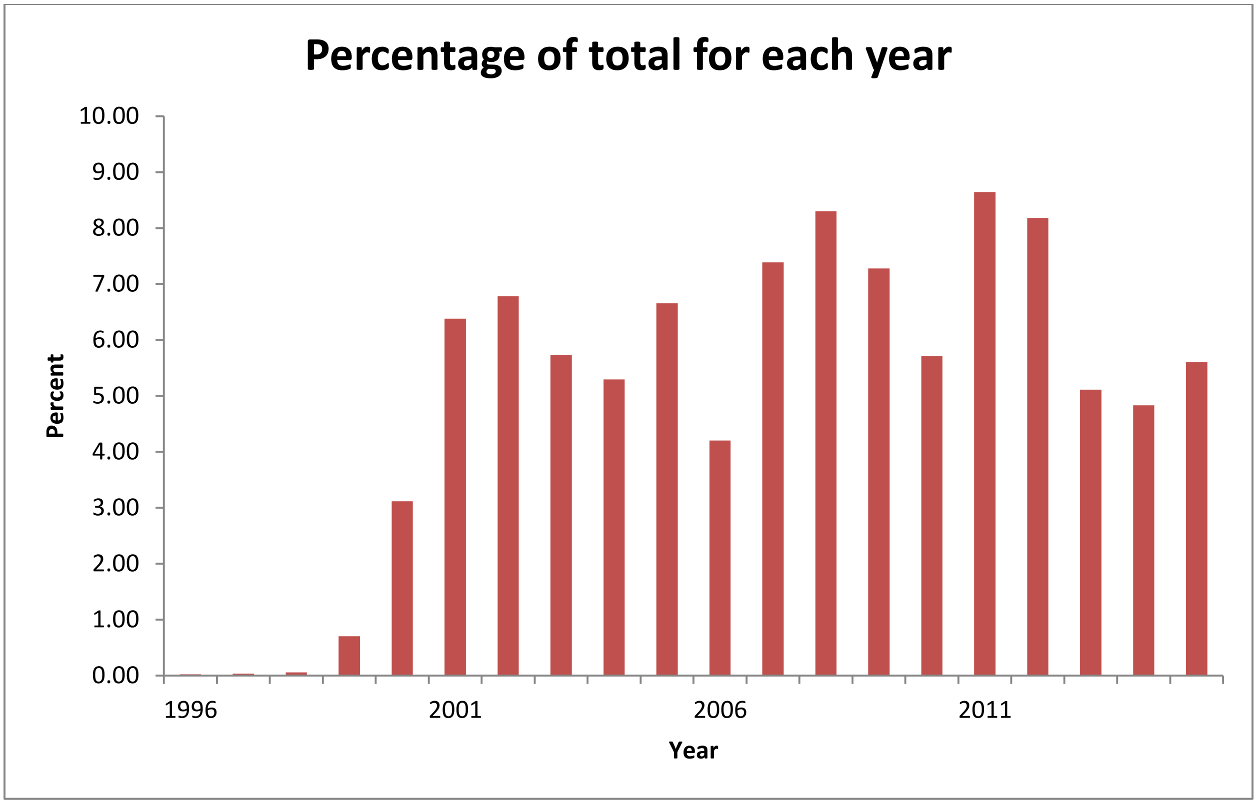

Figure 44: This graph shows the percentage of the total number of entries for each year for the Frogwatch_observation table.

There were a total of 22174 entries for this table at the date of extraction. As we can clearly see the Frogwatch data does not go further back than 1996, and the number of entries was insignificant until 1999.

There is clearly yearly variation of the total number of entries, however it is notable that the peak occurs in 2011, the year NatureWatch was dropped (possibly due to last-minute promotion) and this peak is maintained roughly for 2012, with a sharp drop to 2013 levels, where it remains. What is interesting to note is that despite the sudden drop, it does not seem like the database's participants were completely deserting the program. In fact it may be that the 2011-2012 peak was the result of unusually strong promotion and that the values starting in 2013 more accurately reflect the overall trend.

It must be noted, as well, that as of march 14th there have been no entries for the year 2016. This relates to a point made earlier in the literature review about the requirement for observing frogs as far as citizen science is concerned: the frogs must be “up and about”. Frogs are mostly hidden, hibernating or otherwise dormant, throughout winter.

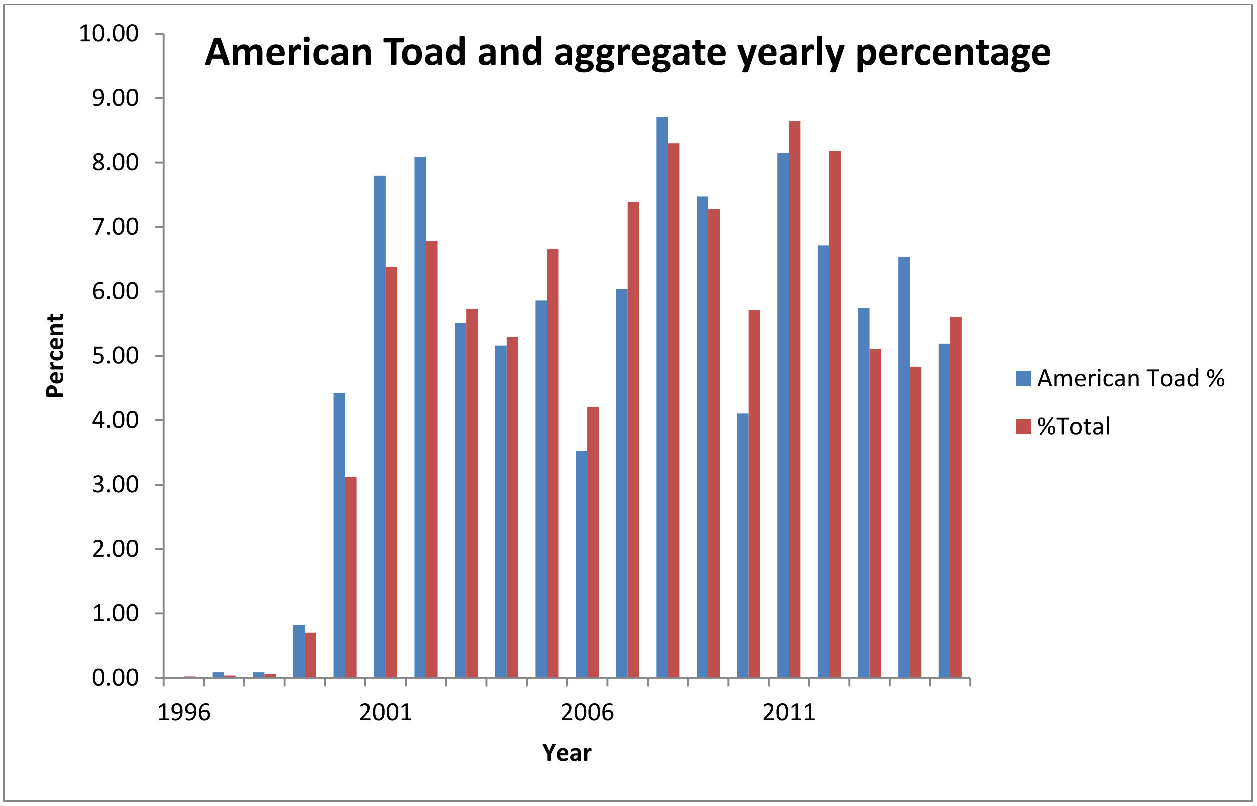

Figure 45: Percentage of entries each year for the American Toad and the aggregate.

There are a total of 3412 dated entries for this species. This makes it usable in analysis.

As we can this from this graph, the largest difference between the American Toad values and the aggregate is that the 2001 and 2002 peaks are much greater for the American toad. The post-2011 drop in numbers is also much less marked than for the aggregate.

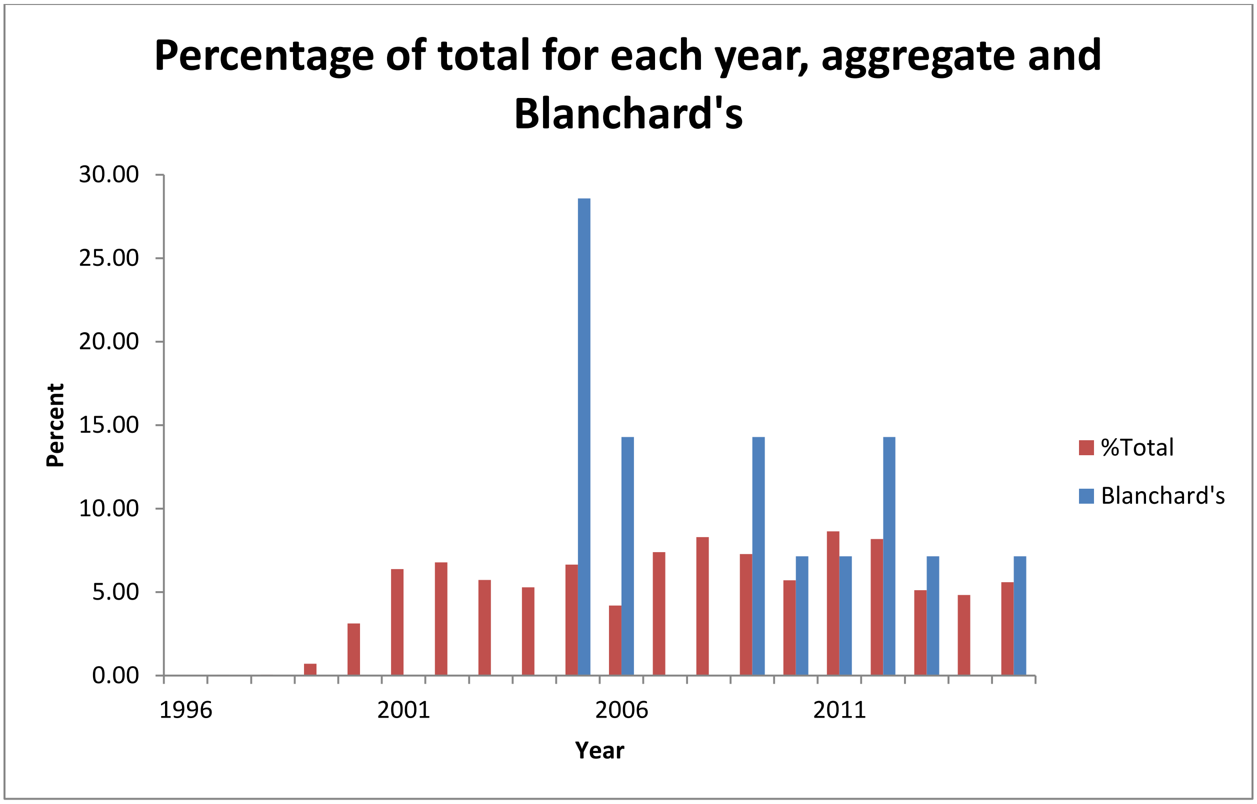

Figure 46: Yearly percentage of the total number of entries for the aggregate and those for Blanchard's cricket frog.

This species is of limited use in analysis as it only has 14 entries with a valid year. The sample size is much too small.

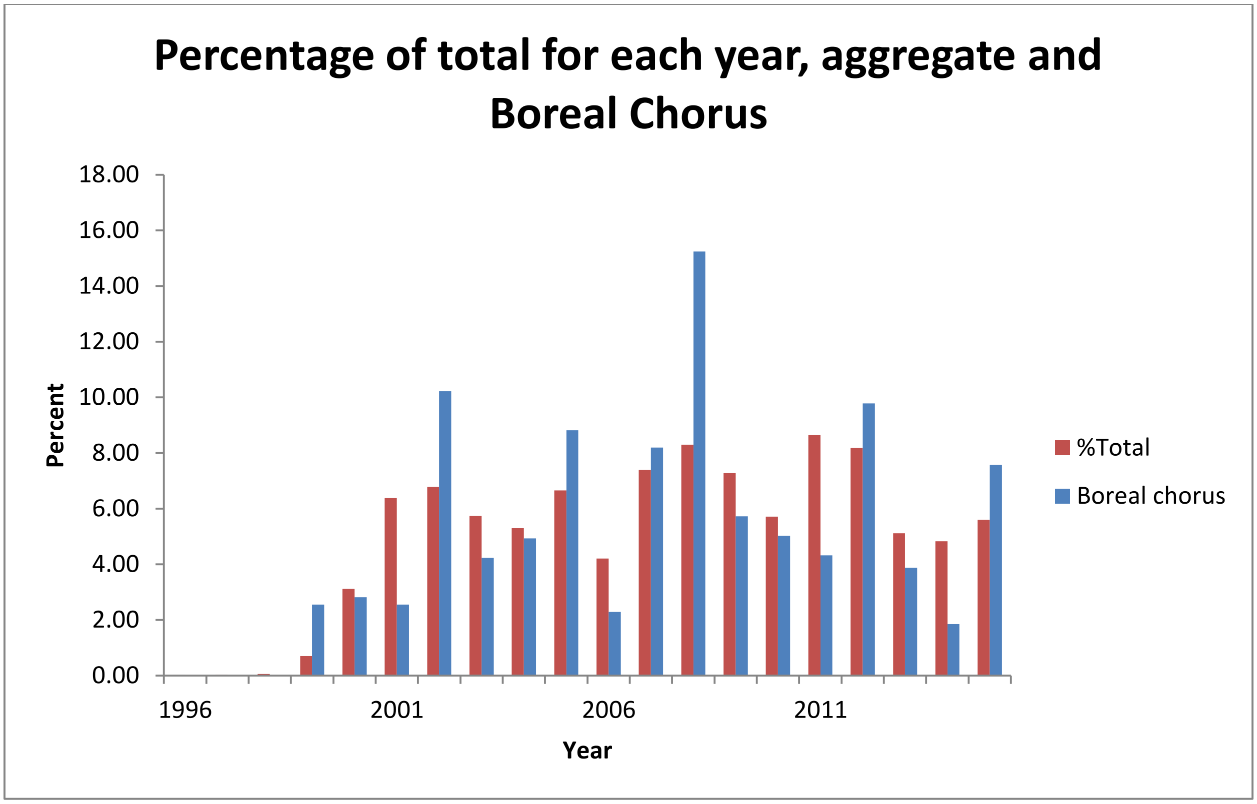

Figure 47: Yearly percentage of the total number of entries for the aggregate and those for the boreal chorus frog.

There were a total of 1135 entries with usable dates for this species.

We can see a number of highly unusual years for this species, notably 2002 and 2008 (high above the aggregate) and 2001 and 2011 (far below the aggregate). In general this species has much greater annual variation than the aggregate, with higher peaks and lower minimum values.

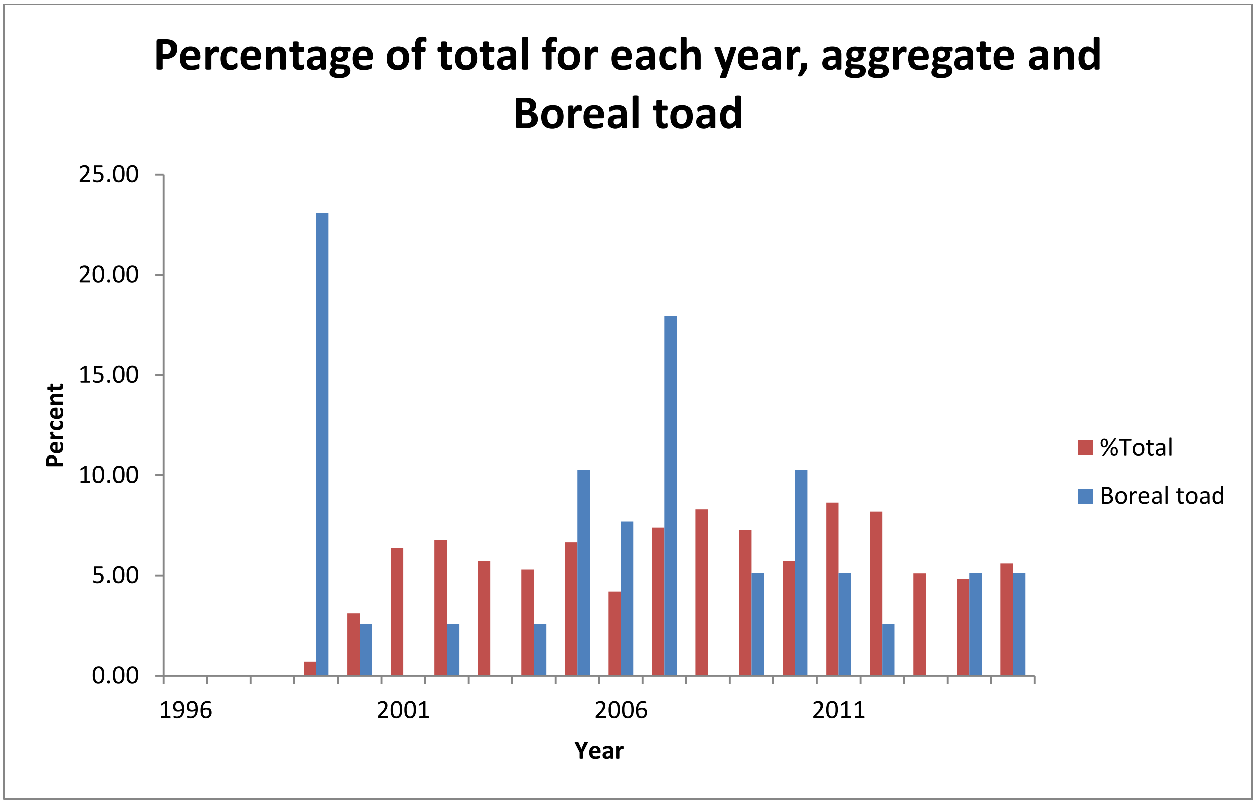

Figure 48: Yearly percentage of the total number of entries for the aggregate and those for the boreal toad.

There are a total of 39 entries for this species, which is too little to make a substantial analysis.

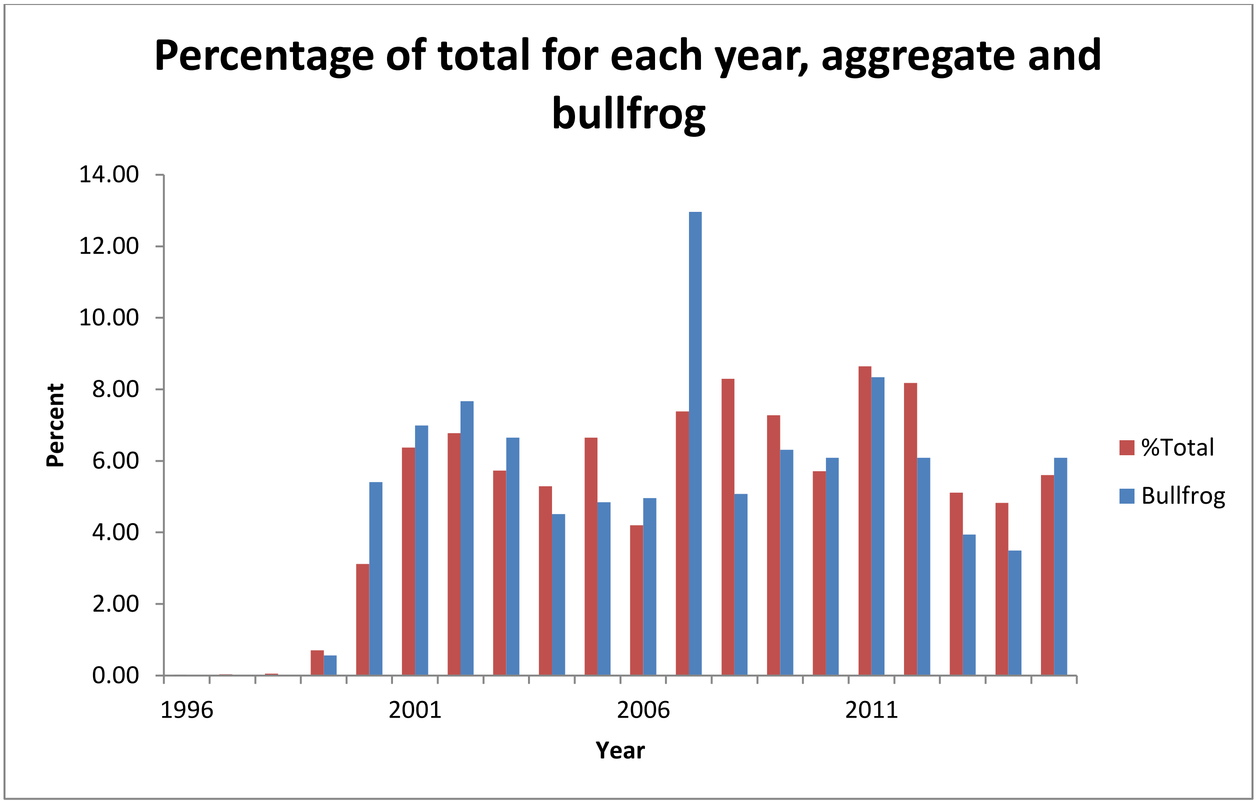

Figure 49: Yearly percentage for the Bullfrog and aggregate.

The total number of usable entries for this species was 887.

The Bullfrog entry numbers largely matches up to the aggregate, except for 2007 where the proportion is much higher. The conditions of this year would be of particular interest to determine why such a boom in observations occurred.

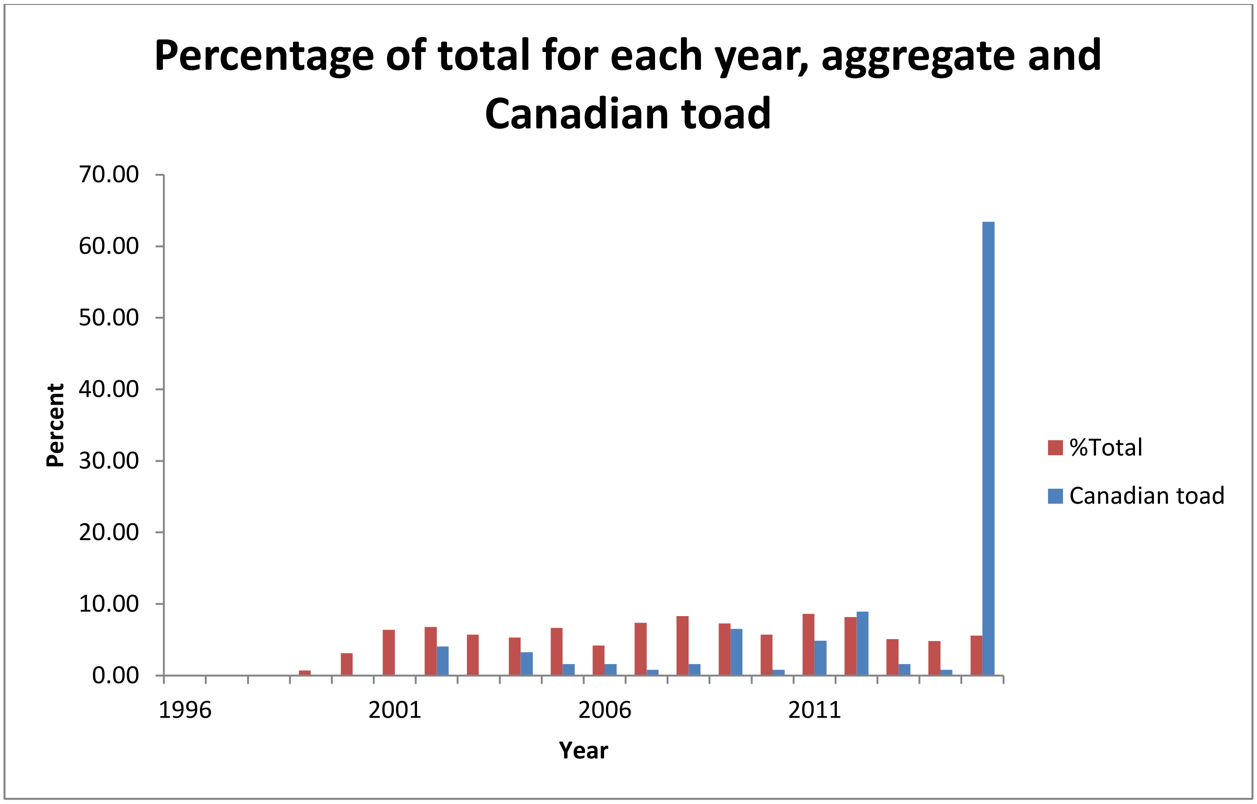

Figure 50: Canadian toad and aggregate yearly percentage.

There were only 123 observations for this species, which is too little to draw conclusions from. The spike in 2015 must be assessed as it may correlate to the massive influx of new observers which occurred late in this year, and which may have put greater emphasis on this species than previously. If this trend of massively greater numbers continues it may reflect on the need to provide detailed information to identify species and make them widely known otherwise they may not be correctly identified nor looked for.

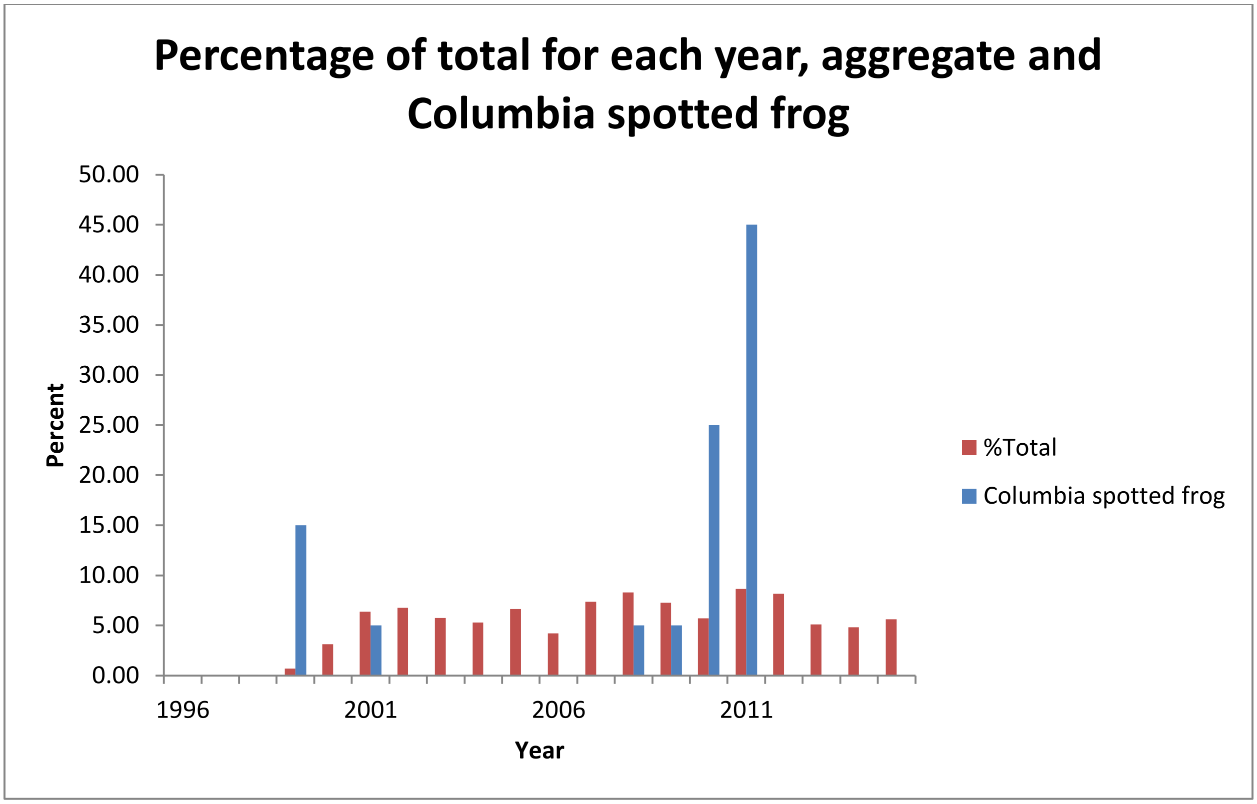

Figure 51: Yearly percentage for the Columbia spotted frog and aggregate.

There are 20 entries for this species, which is too small a sample.

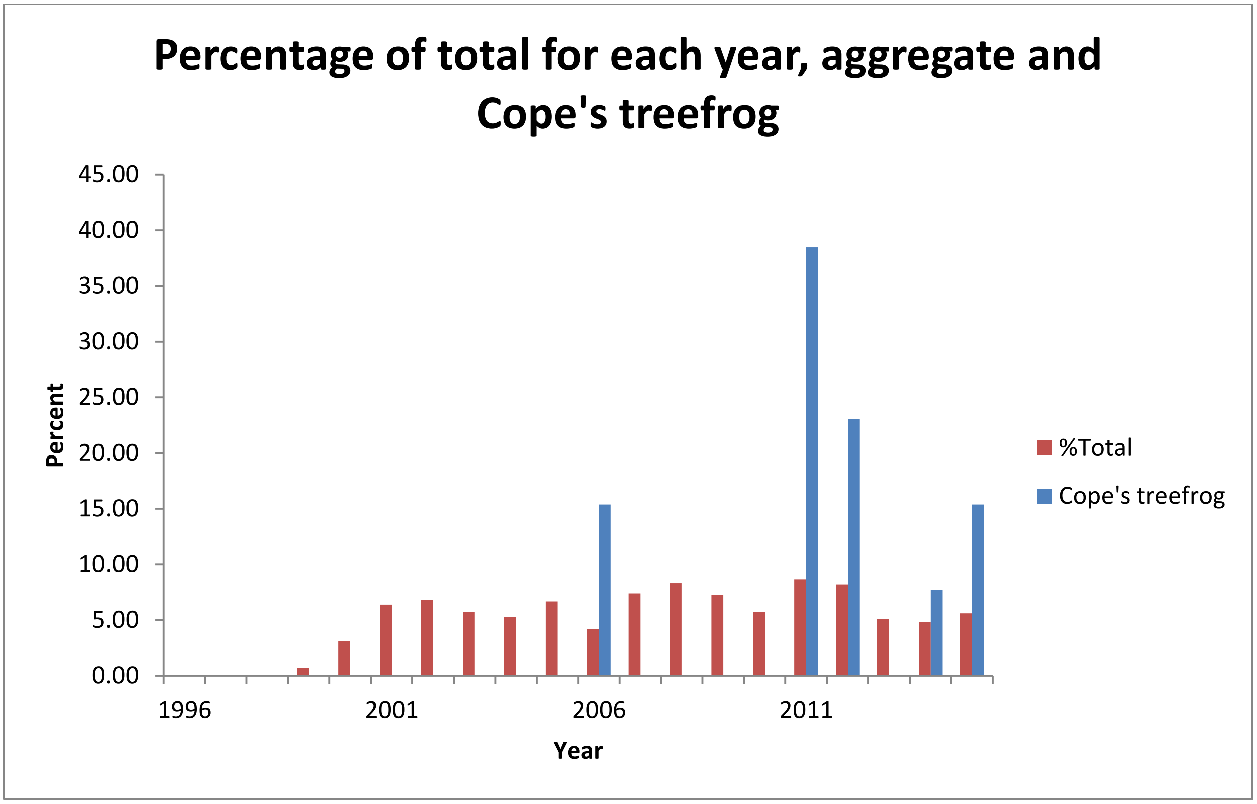

Figure 52: Cope's treefrog and aggregate yearly aggregate.

There are only 13 entries for this species, which is far too little.

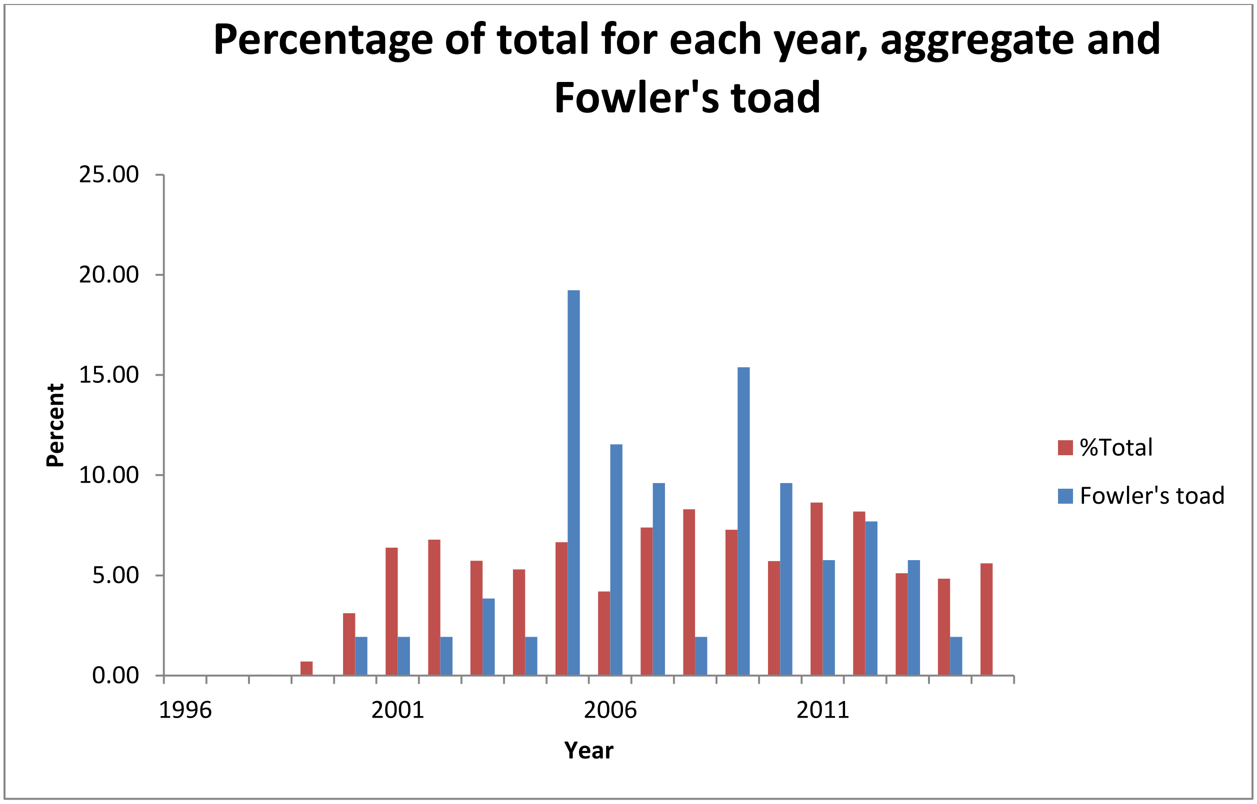

Figure 53: Yearly percentages for Fowler's toad and aggregate.

There are only 52 entries for this species, which is not enough to run conclusive analyses.

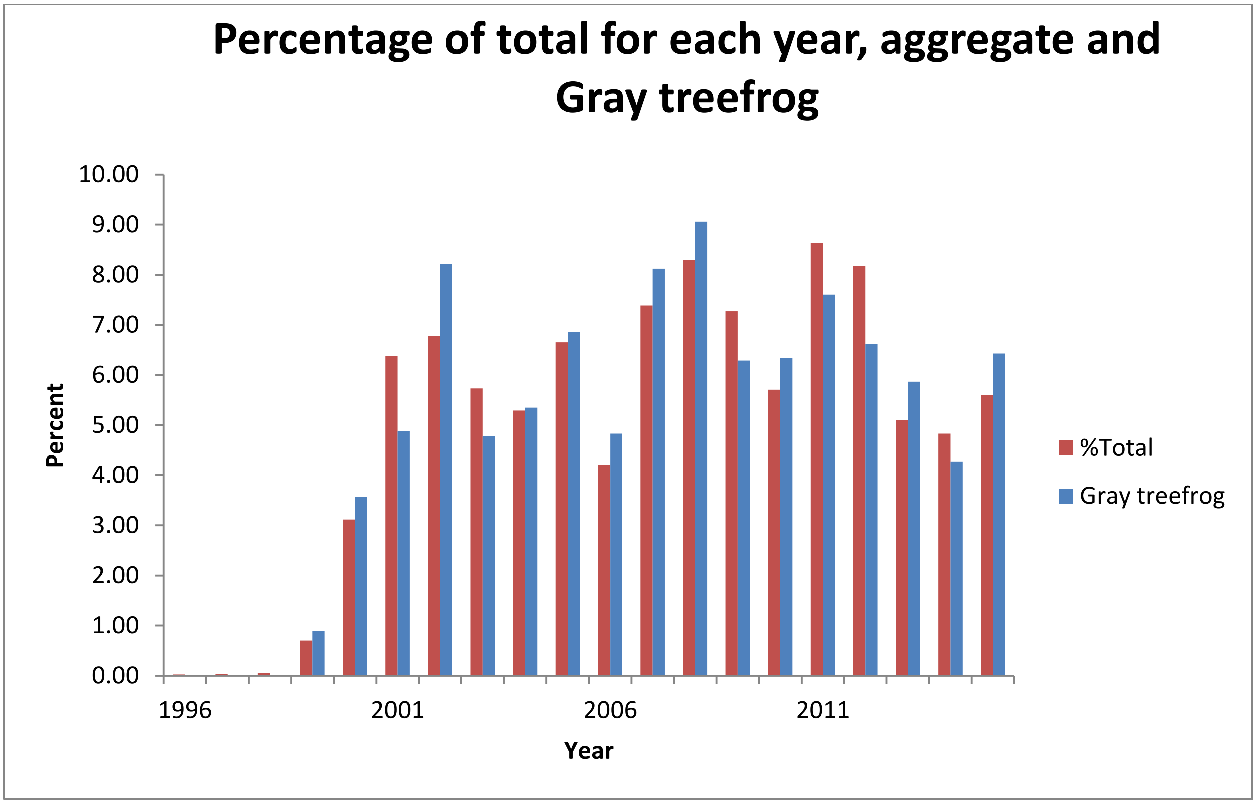

Figure 54: Yearly percentages for the Gray treefrog and aggregate.

There are a total of 2130 entries for this species.

This species follows the aggregate closely for most years, except for 2001, 2002 and 2012. Even then determining if that difference is statistically significant will require further analysis.

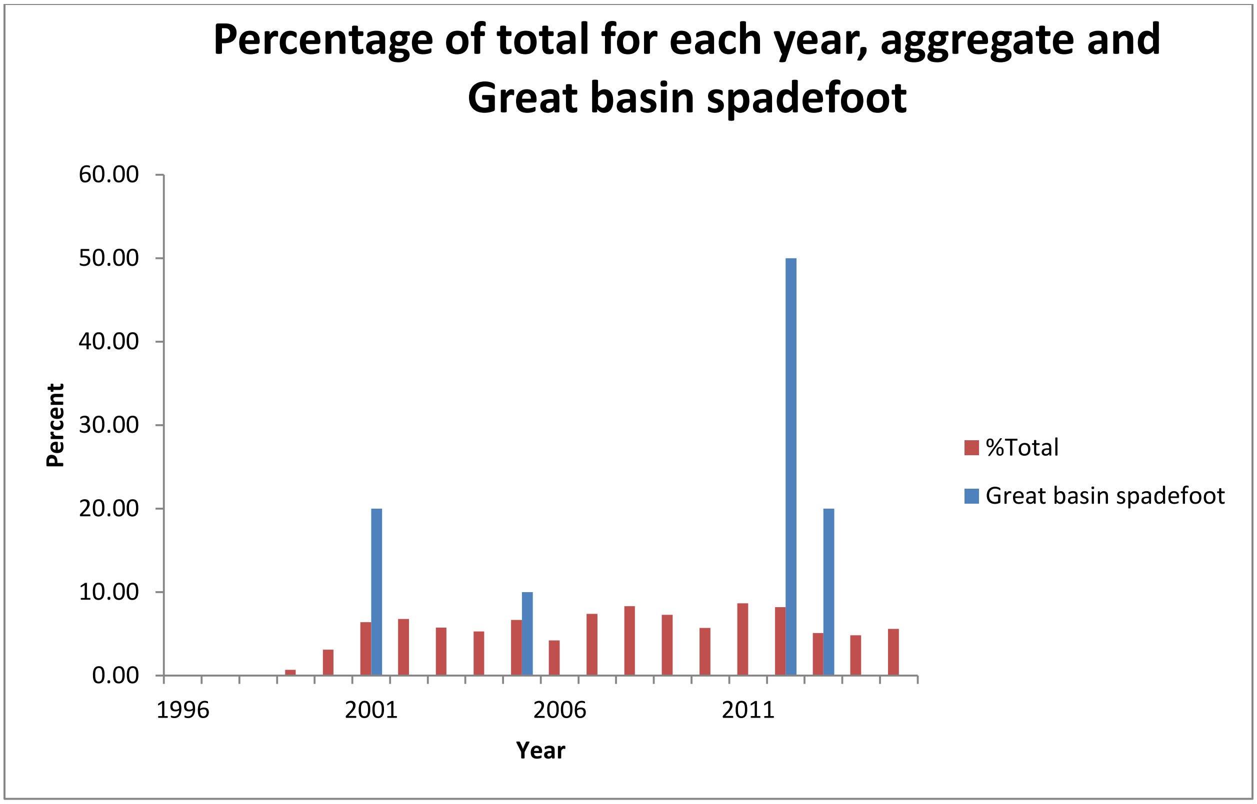

Figure 55: Great Basin spadefoot and aggregate yearly percentages.

There are only 10 entries for this species.

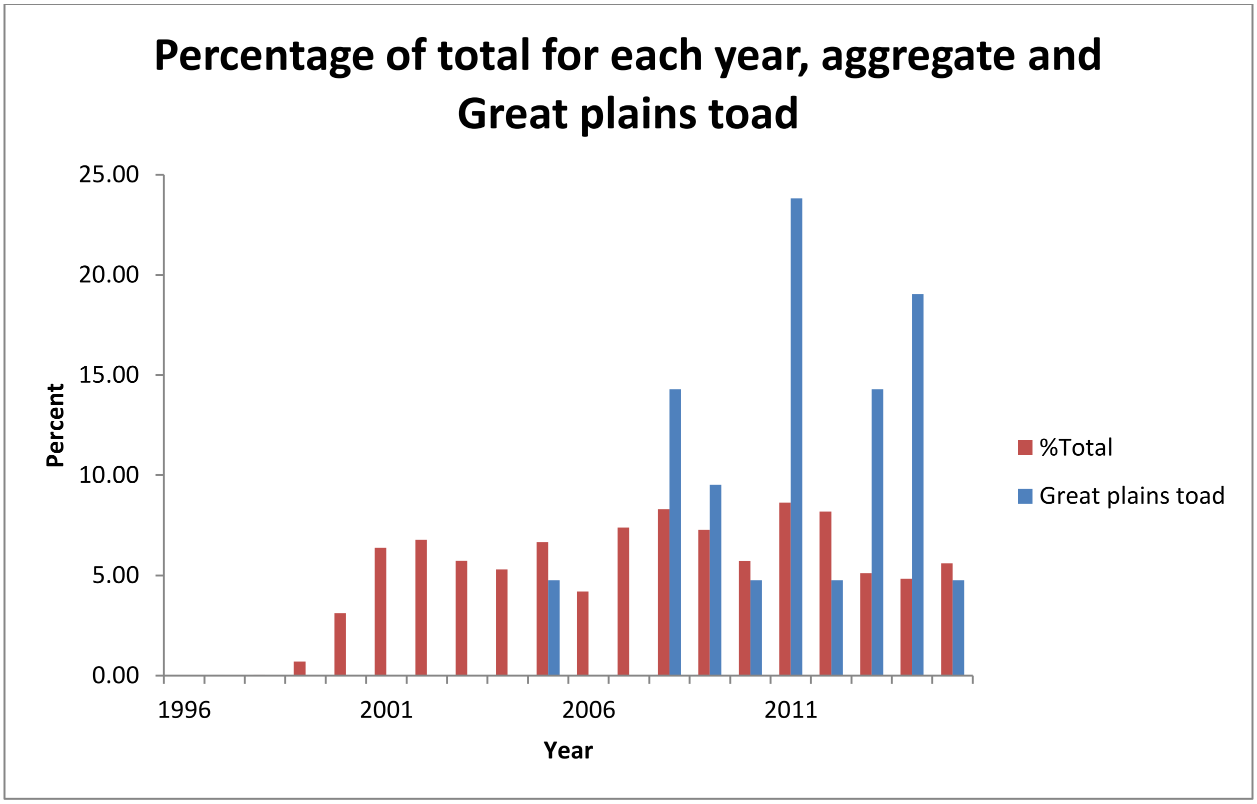

Figure 56: Yearly percentages for the Great Plains toad and aggregate.

There are only 21 entries for this species.

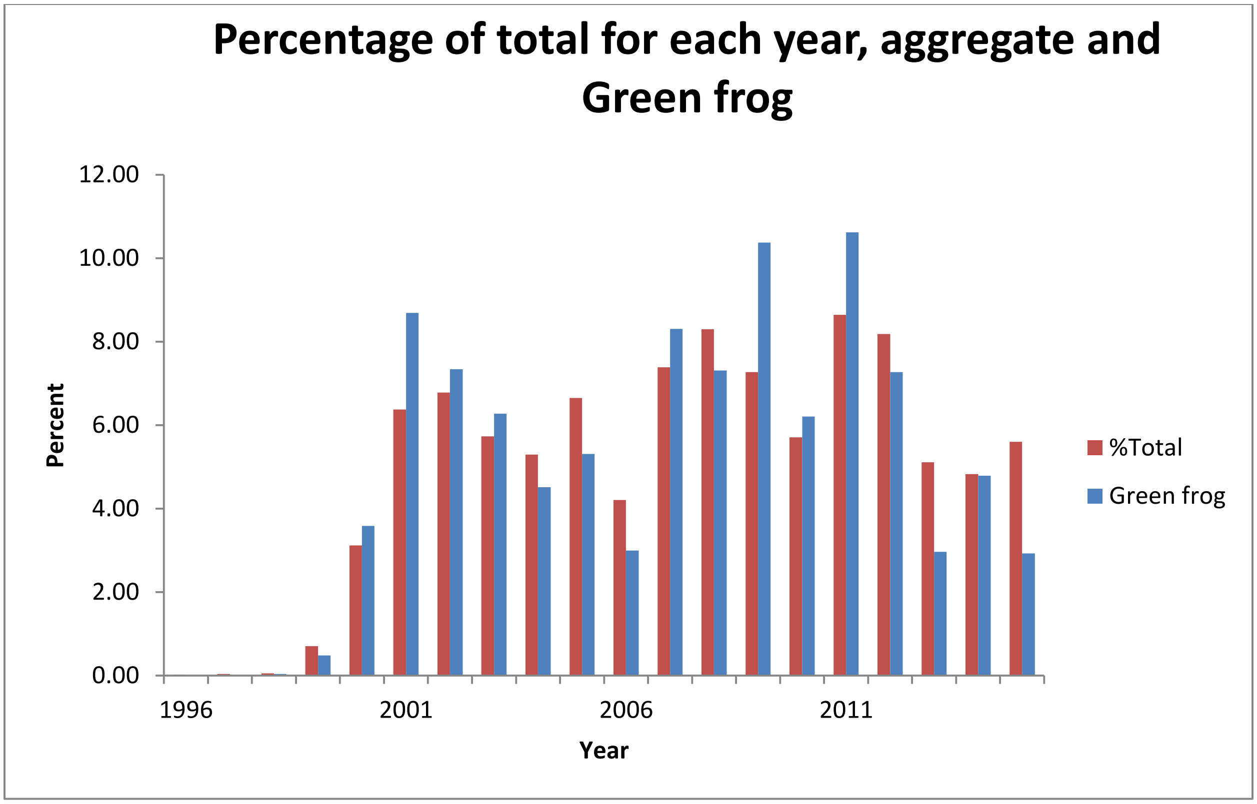

Figure 57: Green frog and aggregate yearly percentages.

There are a total of 2901 entries for this species.

The percentage of entries was much greater than it was for the aggregate for the years of 2001, 2009 and 2011.

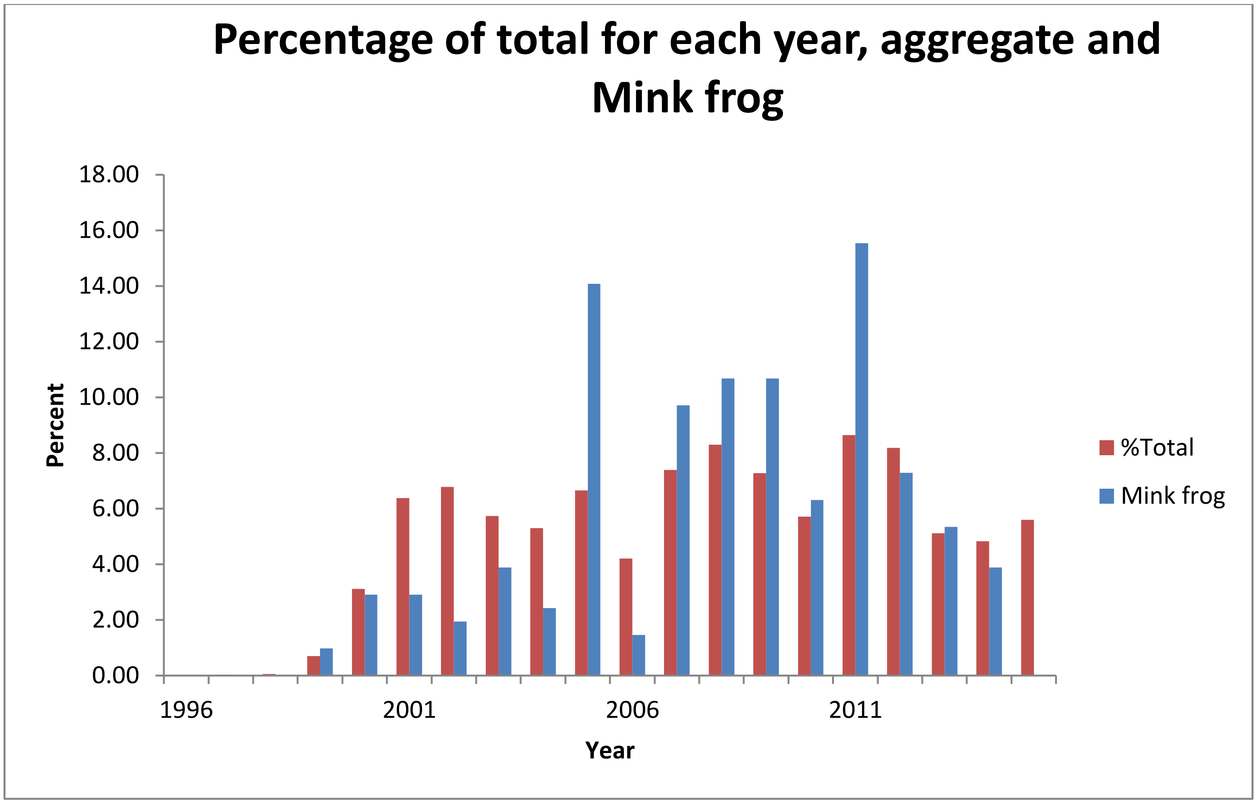

Figure 58: Yearly percentages for the Mink frog and aggregate.

This species has 206 usable entries. This is a little too low to be useful on its own but can be integrated into a larger analysis.

This species has very marked anomalies for the years 2005 and 2011.

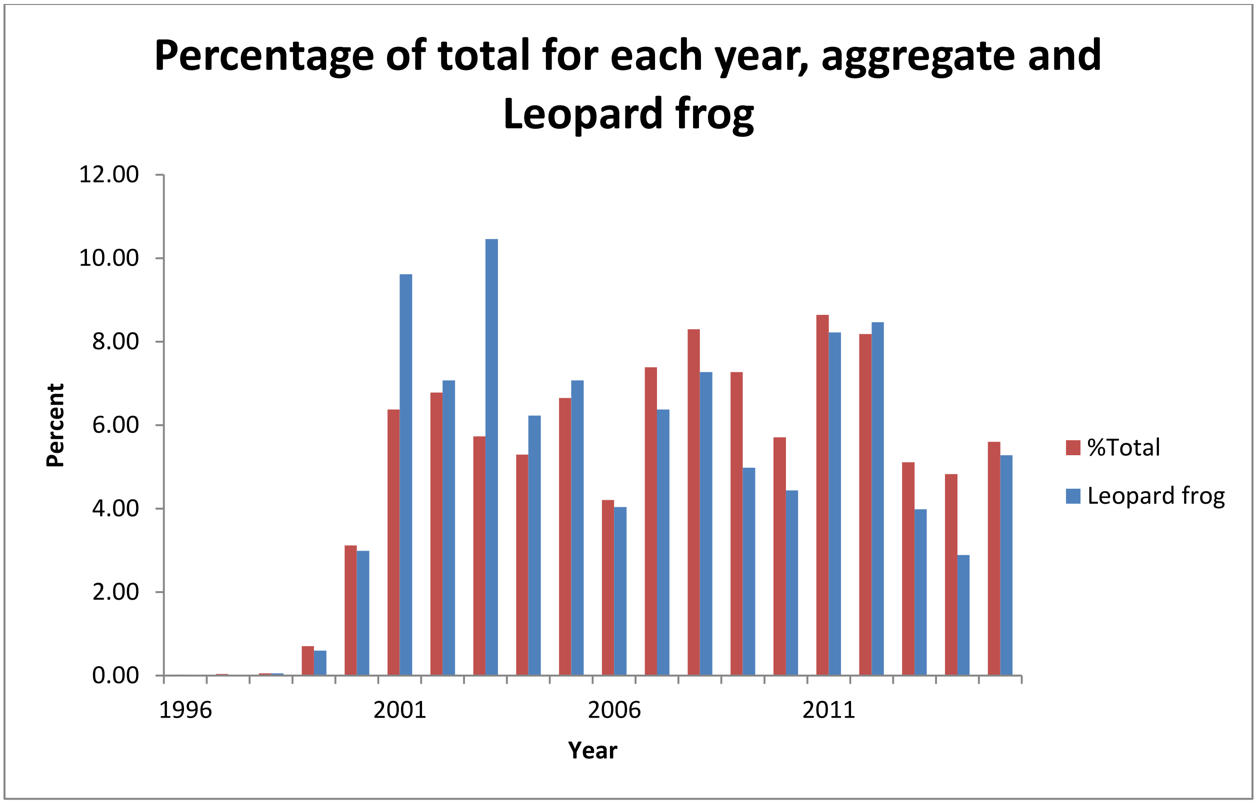

Figure 59: Leopard frog and aggregate yearly percentages.

This species has a total of 2008 entries.

As we can see the main difference is in the abnormally high proportion of Leopard Frog entries in 2001 and 2003.

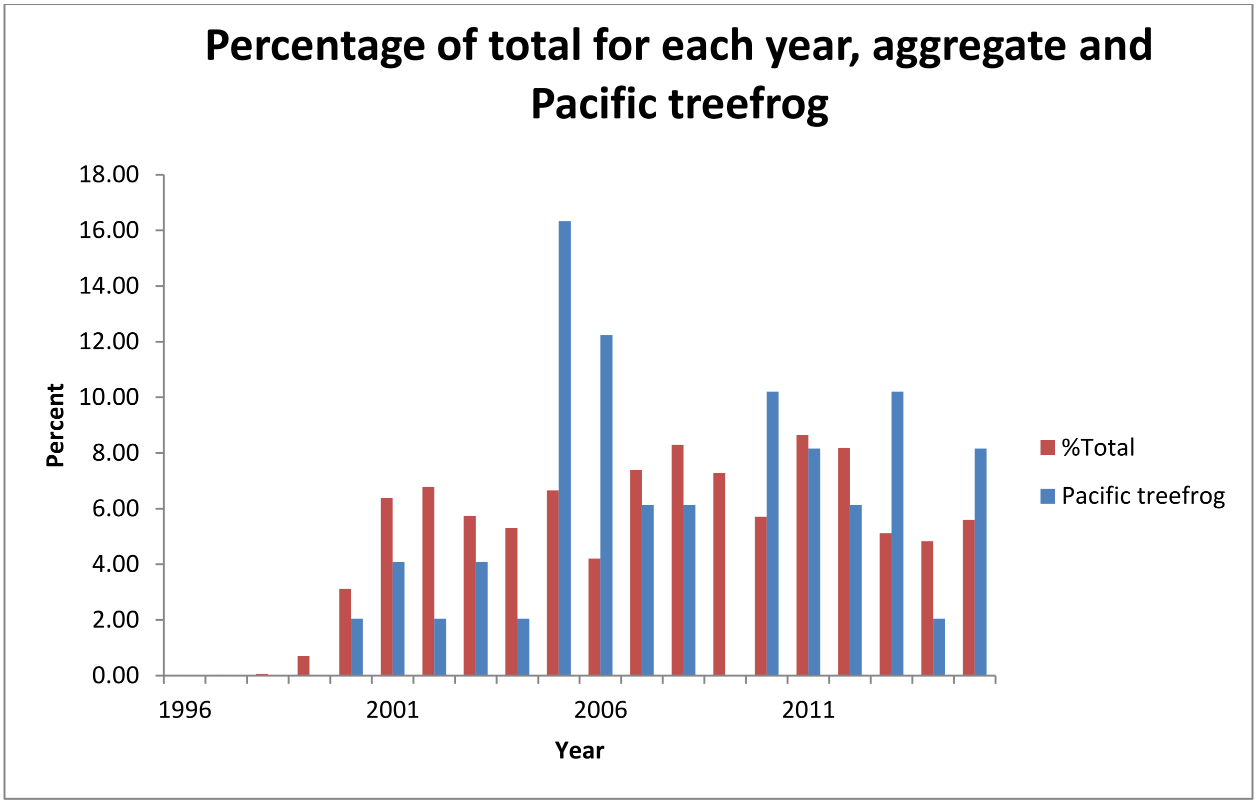

Figure 60: Pacific treefrog and aggregate yearly percentages.

There are only 49 Pacific treefrog entries.

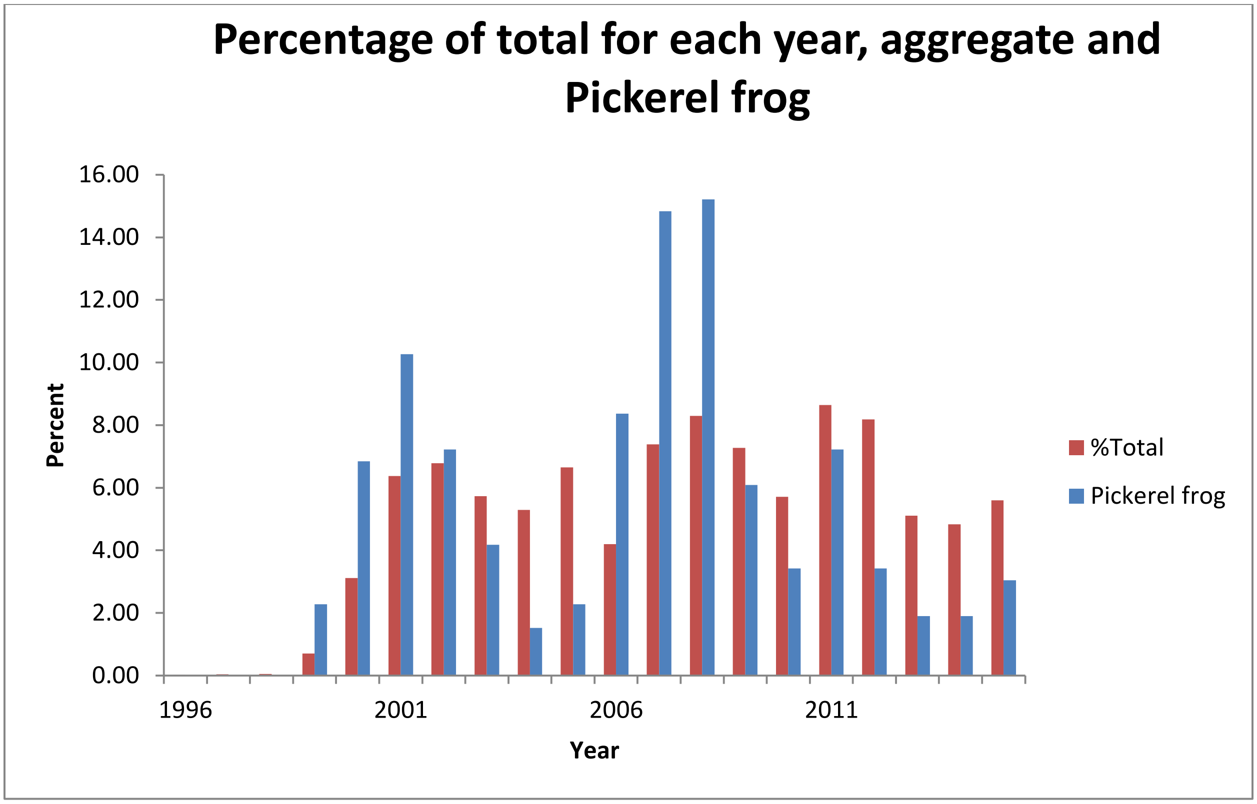

Figure 61: Yearly percentages for the aggregate and Pickerel frogs.

There are 263 Pickerel frog entries.

This is too little to draw conclusions.

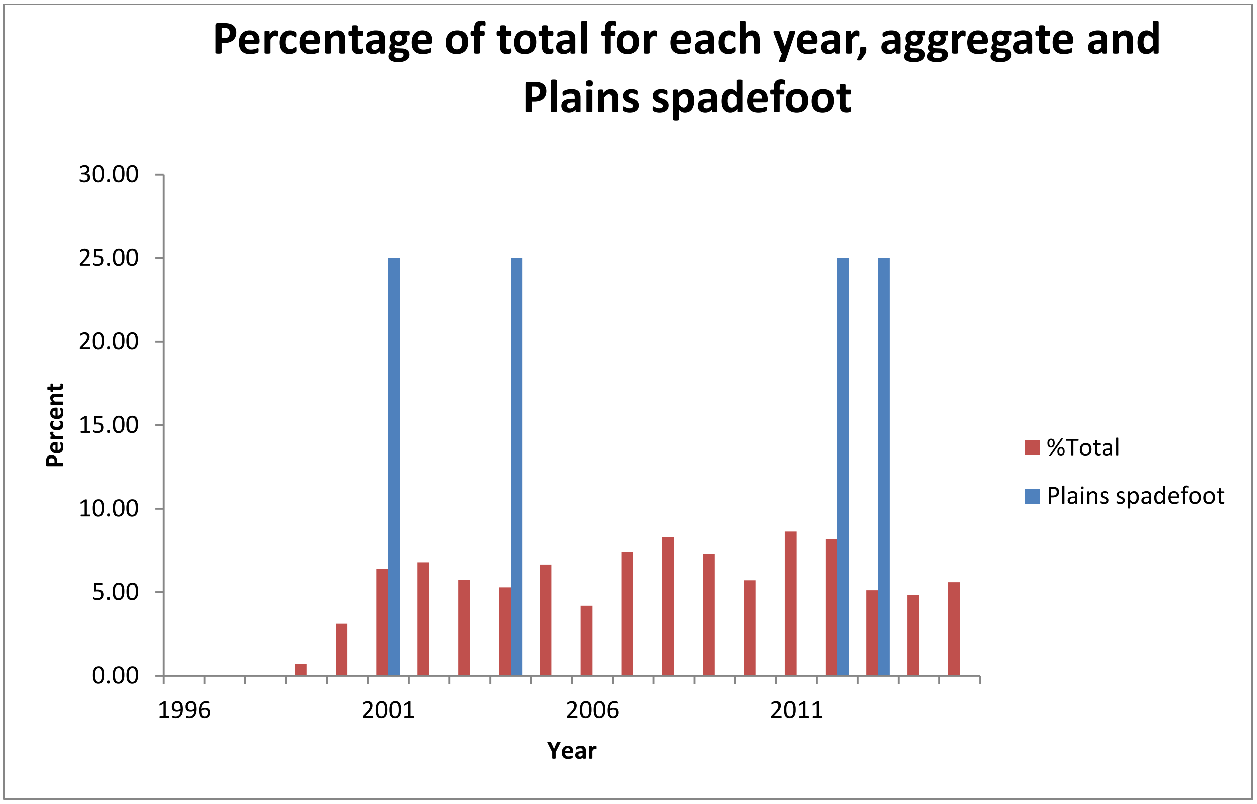

Figure 62: Plains spadefoot and aggregate yearly percentages.

There are only 4 entries for this species.

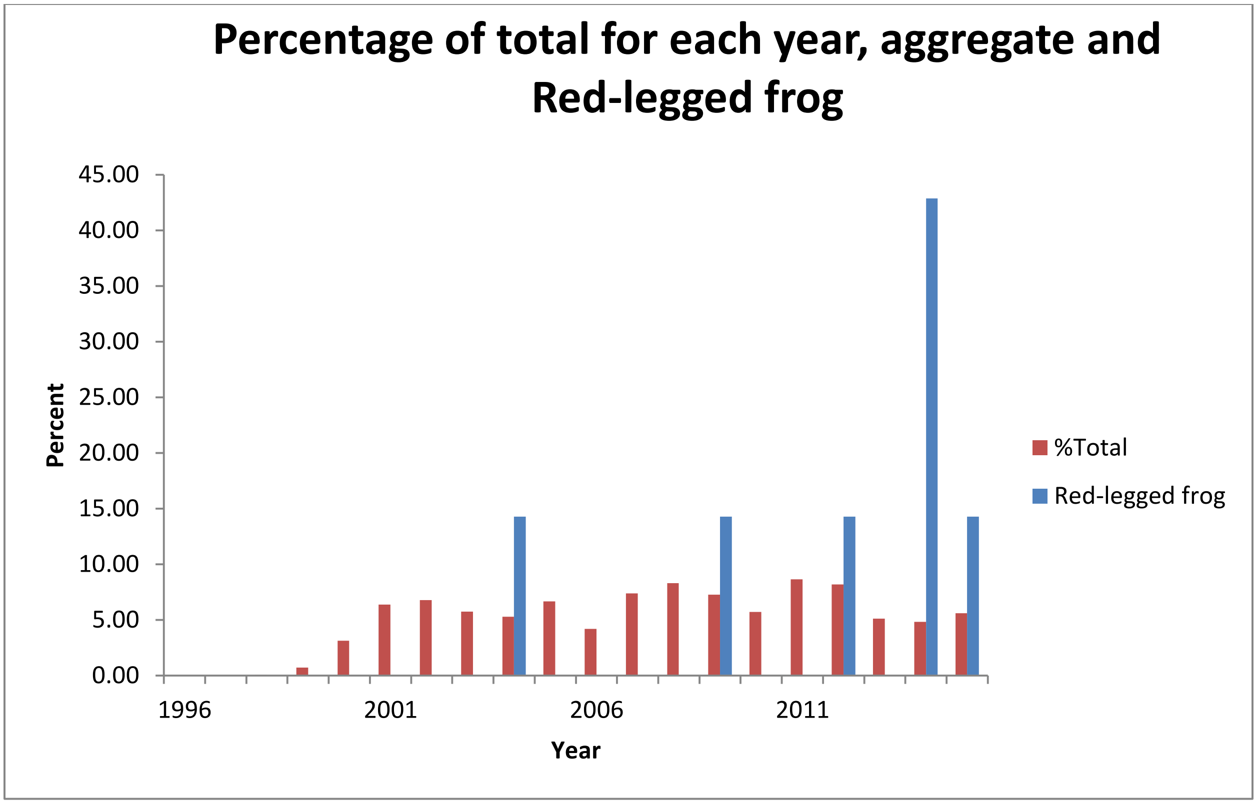

Figure 63: Red-legged frog and aggregate yearly percentages.

There are only 7 entries for this species.

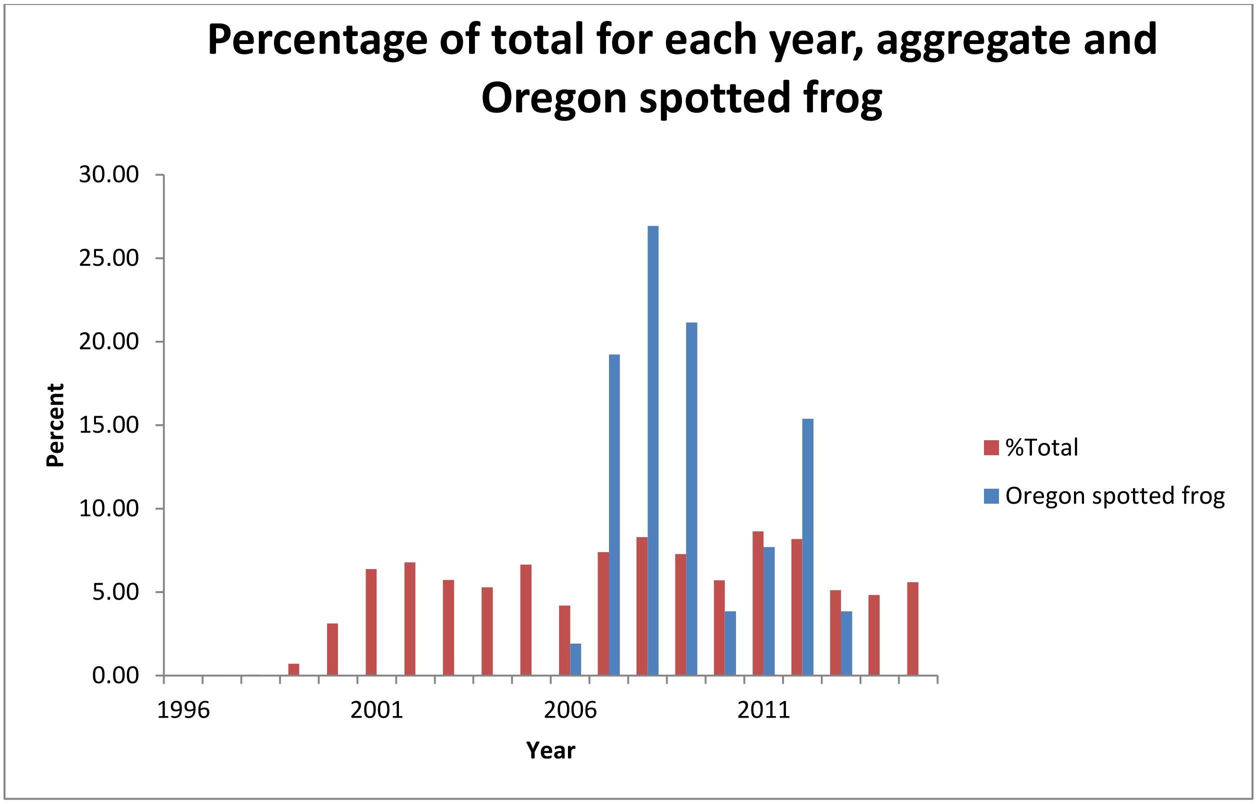

Figure 64: Yearly percentages for Oregon spotted frog and aggregate.

There are only 52 entries for this species.

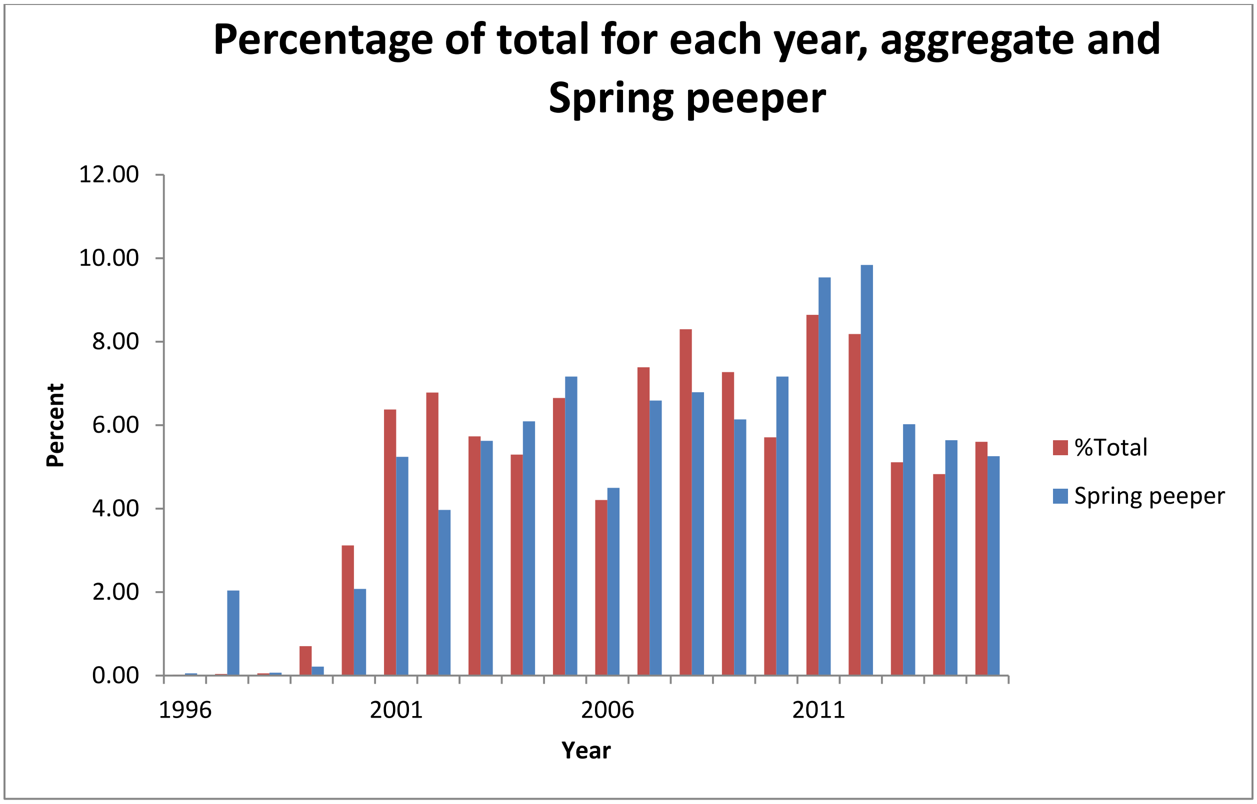

Figure 65: Percentage of total entries by year, for the Spring peeper entries and the aggregate.

This species, with a total entry count of 6028, is the most abundant in the database.

As we can see, other than for the year 2002, the Spring Peeper's numbers largely mirror that of the aggregate.

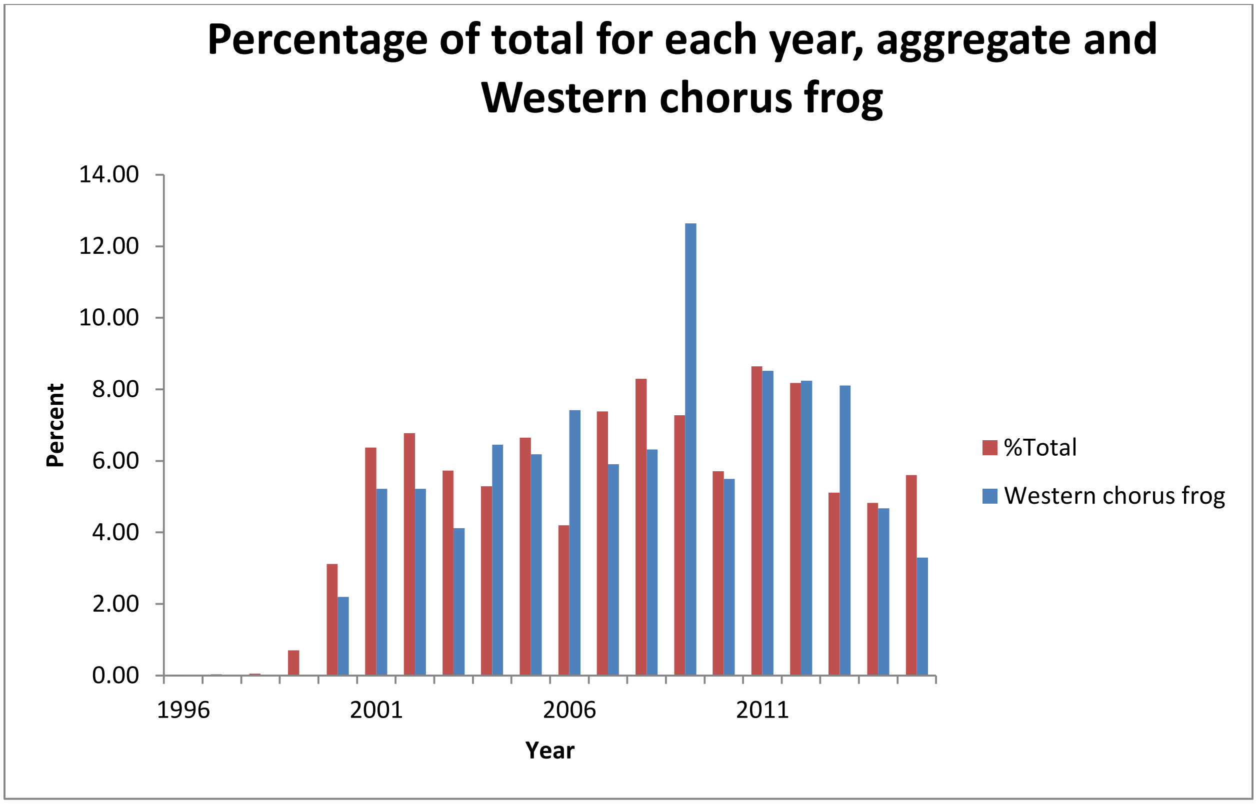

Figure 66: Western chorus frog and aggregate yearly percentages.

There are 728 entries for this species currently.

The 3 most notable trends for this species are for the years 2006, 2009 and 2013. The strongest out-of-pattern event was in 2009, so this year is of considerable interest.

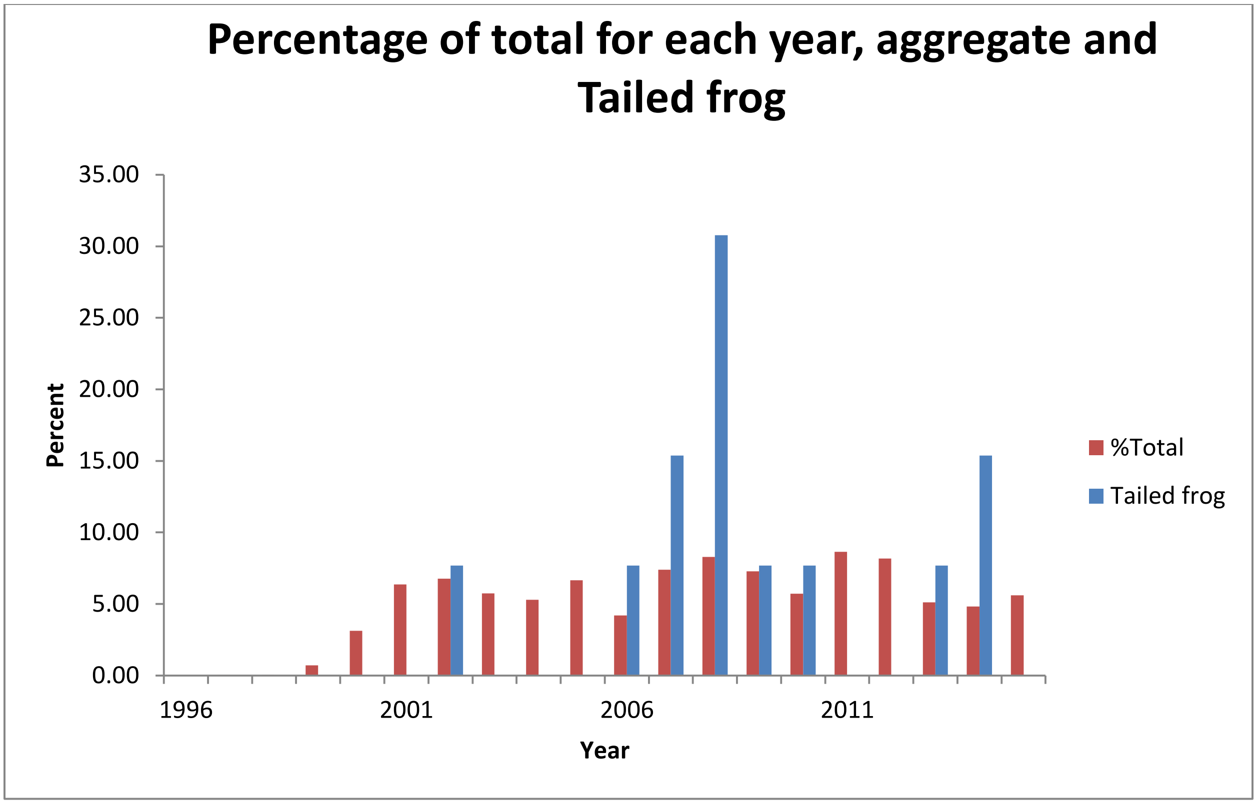

Figure 67: Yearly percentages for the aggregate and Tailed frog.

This species only has 13 usable entries, so we cannot make much use of the data.

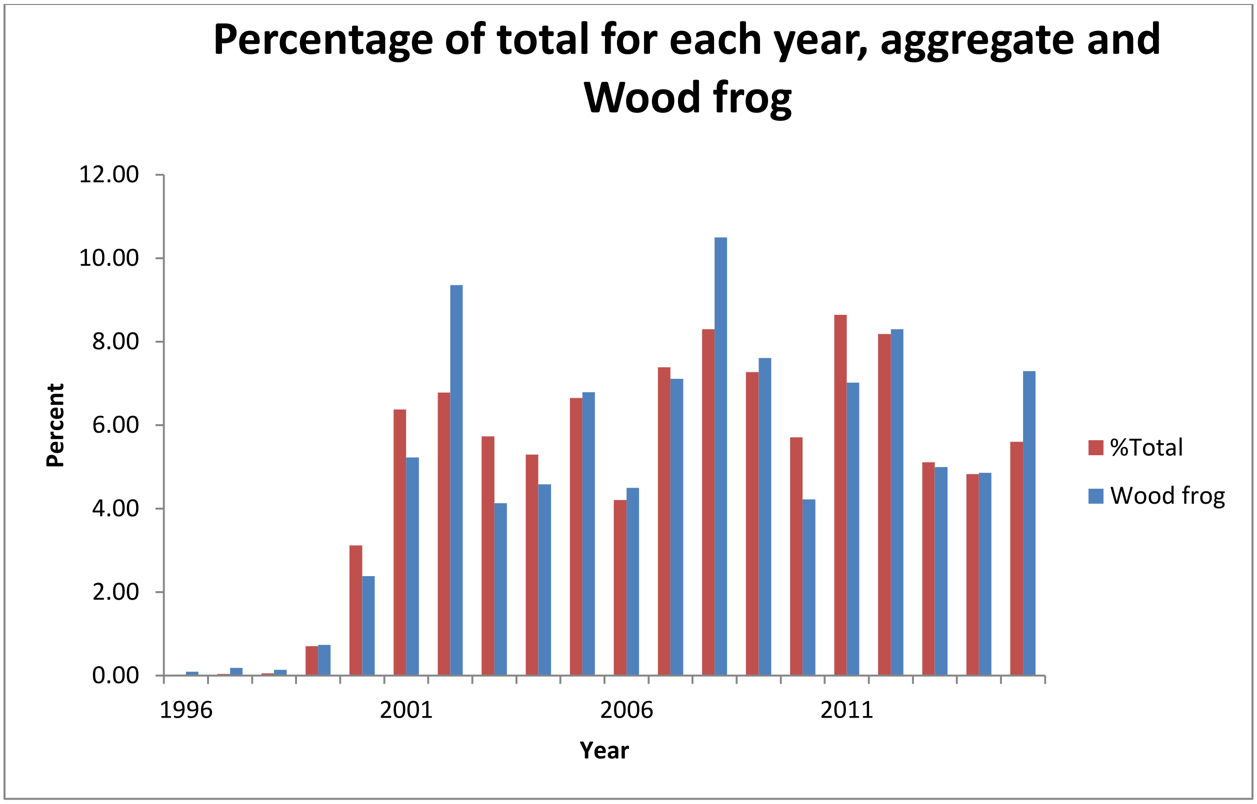

Figure 68: Wood frog and aggregate yearly percentages.

There are 2181 entries for this species, which is a very good number for analysis.

For the most part, like most of the high-entry number species, the values for this species follows the aggregate closely with exceptions for a few years. The most notable being the years of 2002, 2008, and possibly 2015.

Appendix A-3: Frogwatch spatial analysis

This basic analysis was done through ArcGIS' ArcMap program. After exporting the frogwatch_observation data and the Locations data, the locations were assigned XY coordinate data based on the Longitude and Latitude (LONG and LAT) columns. The frogwatch_observation data was then joined to the newly-mapped locations layer (not the other way around otherwise each location will only register a single data point – whereas there is overlap). A layer was then created for each species by the following process: Select by attribute (species number column) – unique value – Create layer from selected features.

This was further refined by removing the "absence" data (where the observation code = 1) from the layers.

With these layers I could then run a point density operation to show roughly were the observations were concentrated. If this were not citizen science data, we could reasonably infer that where these areas of high entry density are correspond most with actual presence – however as it is we must consider the possibility that entry density is heavily influenced by the potential "observer density" i.e. entries show up where people look for them (and fauna in general is obviously not limited to those areas where people often look for them).

The coordinates provided for the locations are in degree decimals, with some 3 decimal spaces. This gives us a precision of 111 meters for latitude and around 80 meters for longitude.

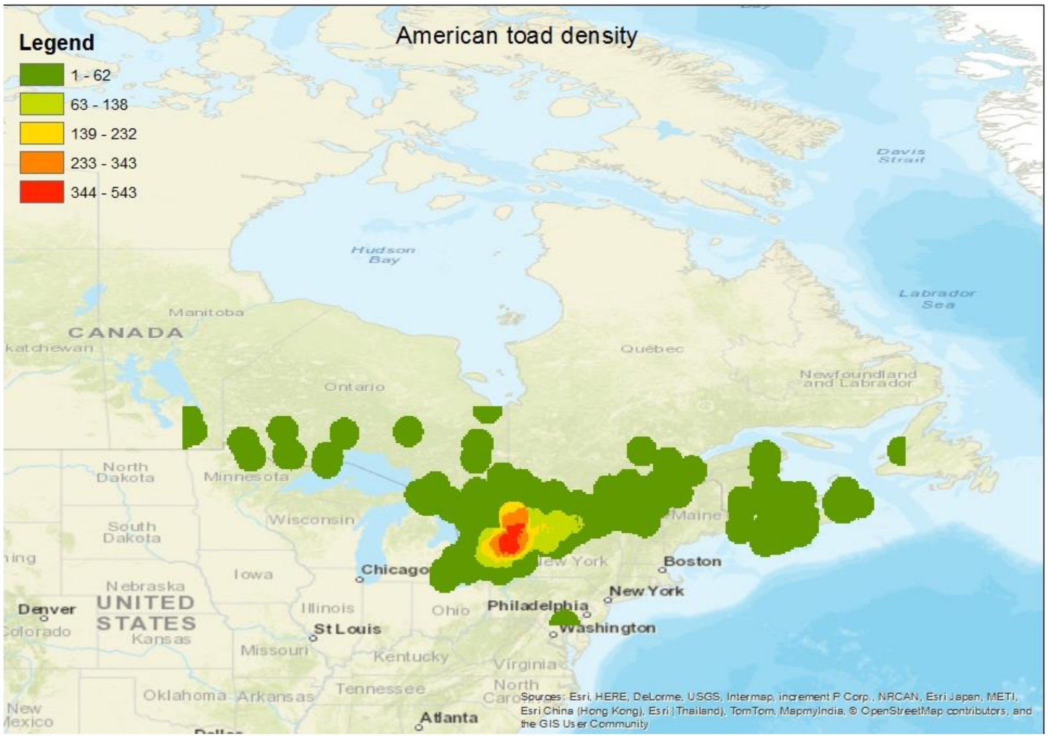

Figure 69: American toad entry density map.

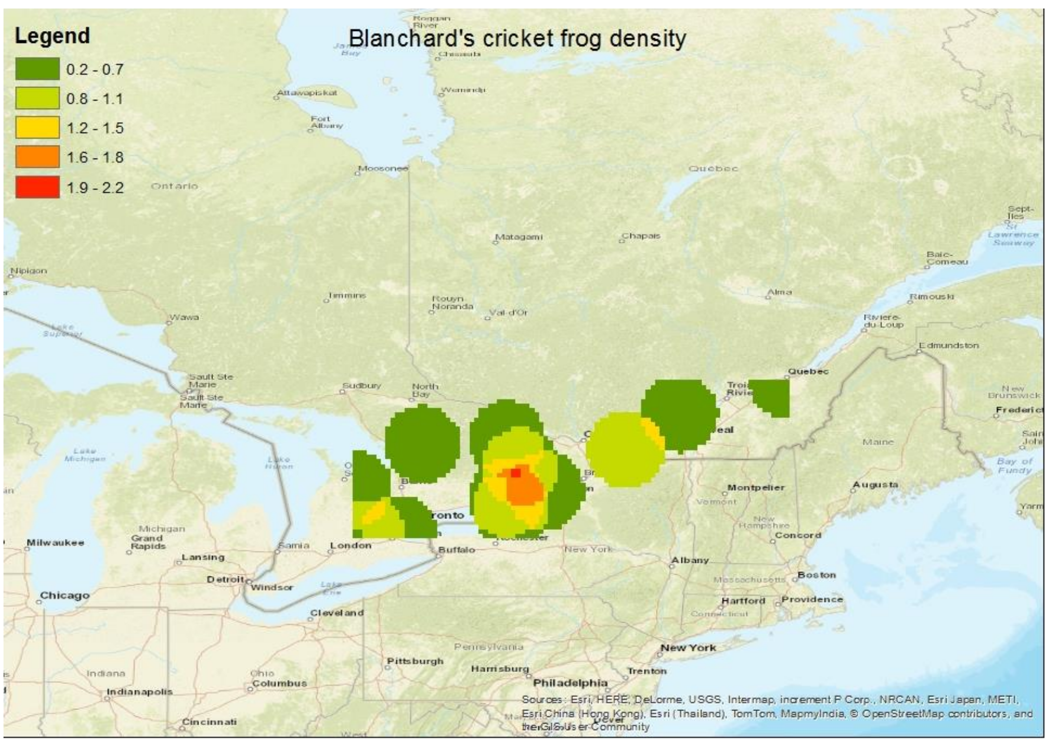

Figure 70: Blanchard's cricket frog entry density map.

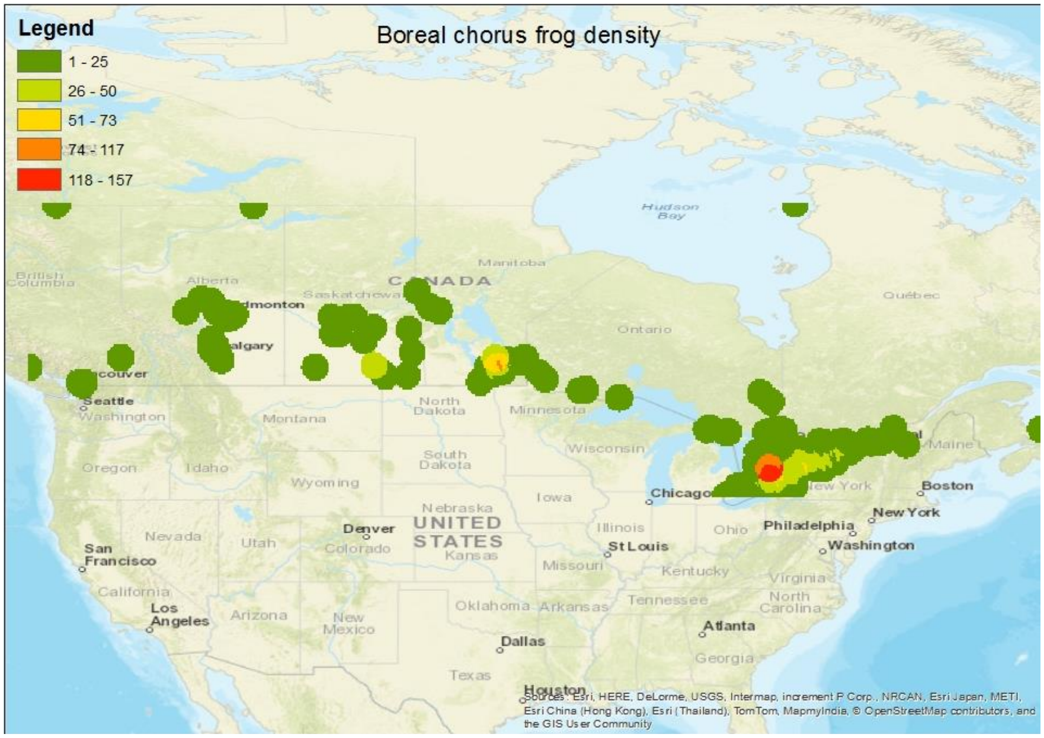

Figure 71: Boreal chorus frog entry density map.

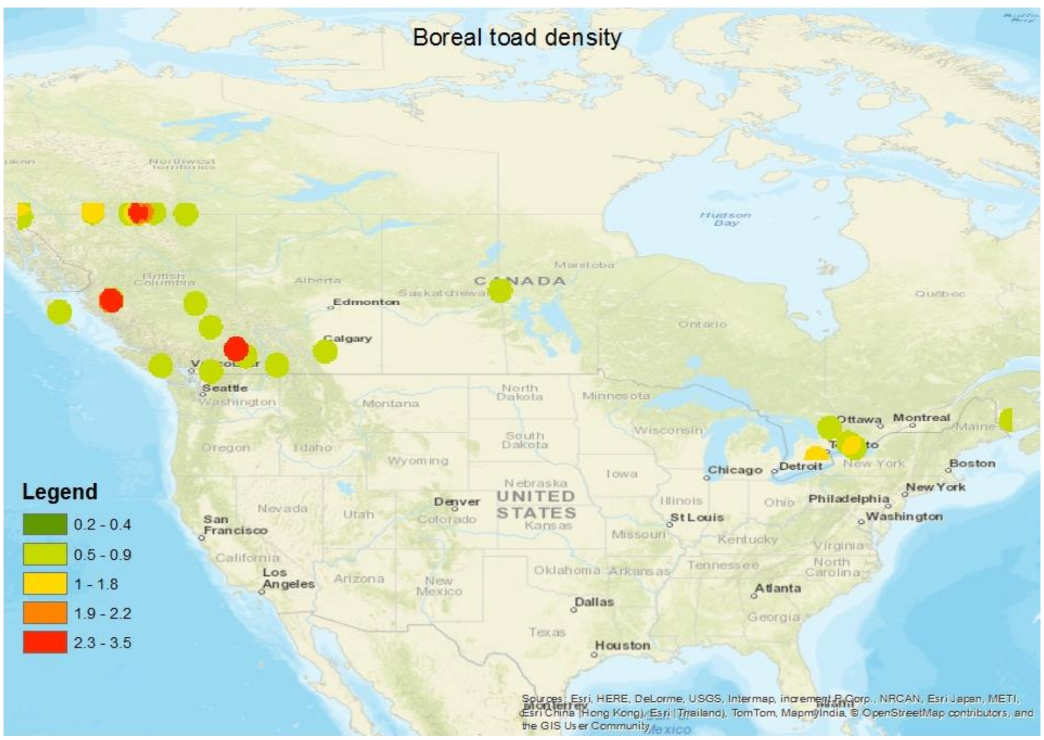

Figure 72: Entry density map for the Boreal toad.

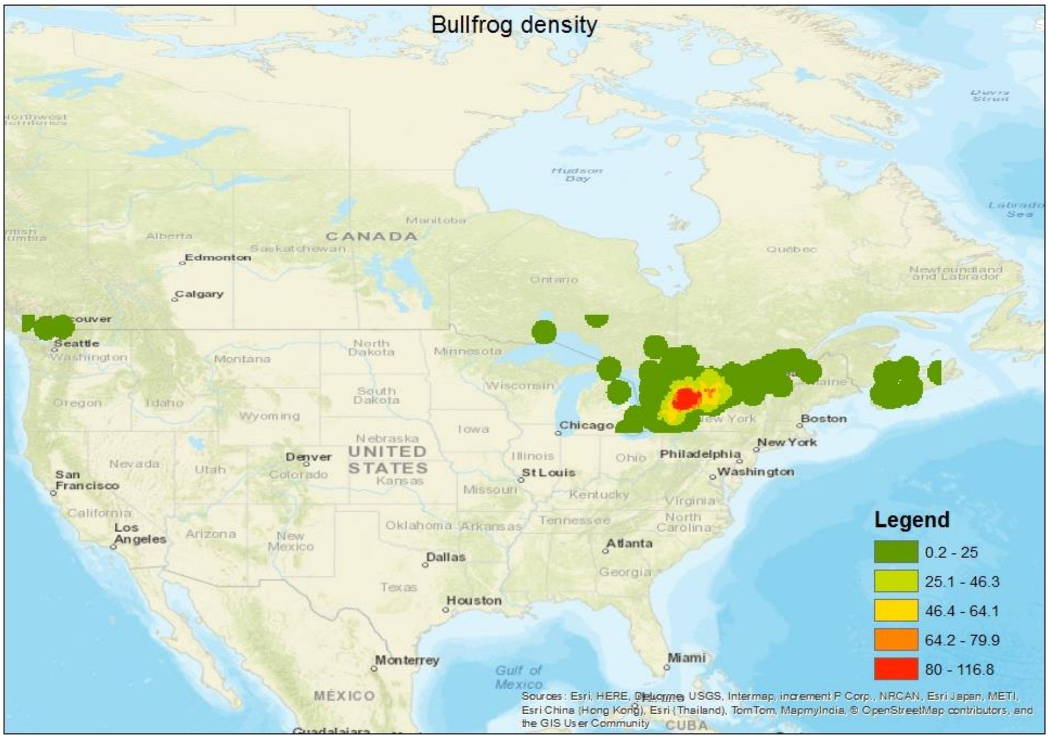

Figure 73: Bullfrog entry density map.

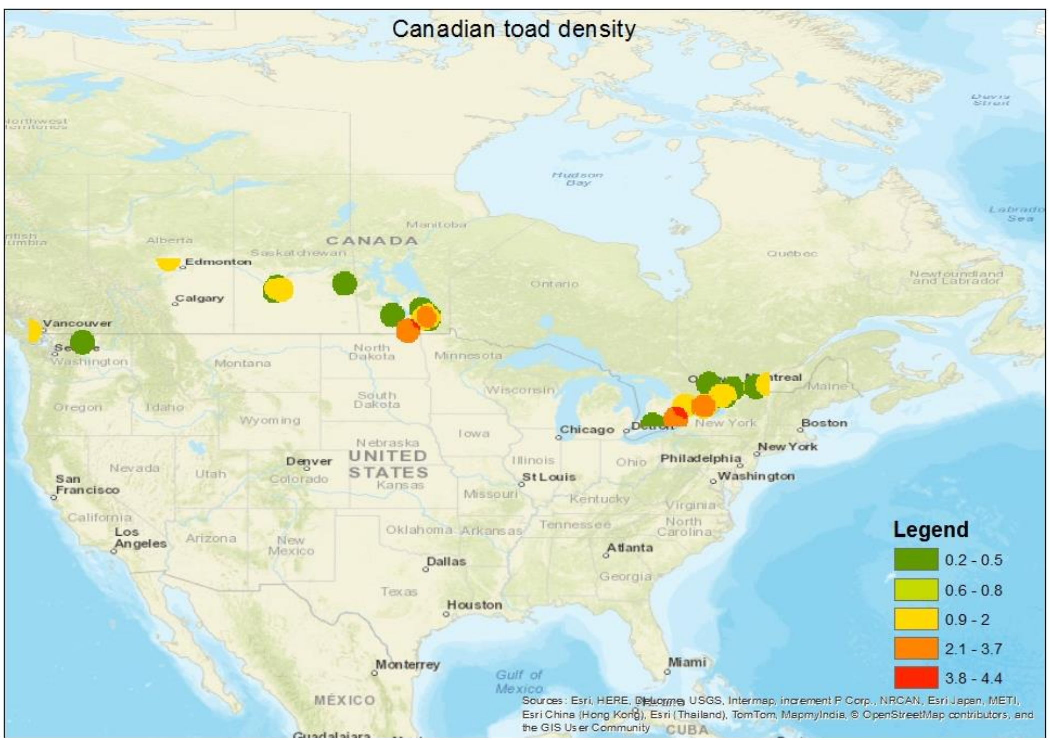

Figure 74: Entry density map for Canadian toad.

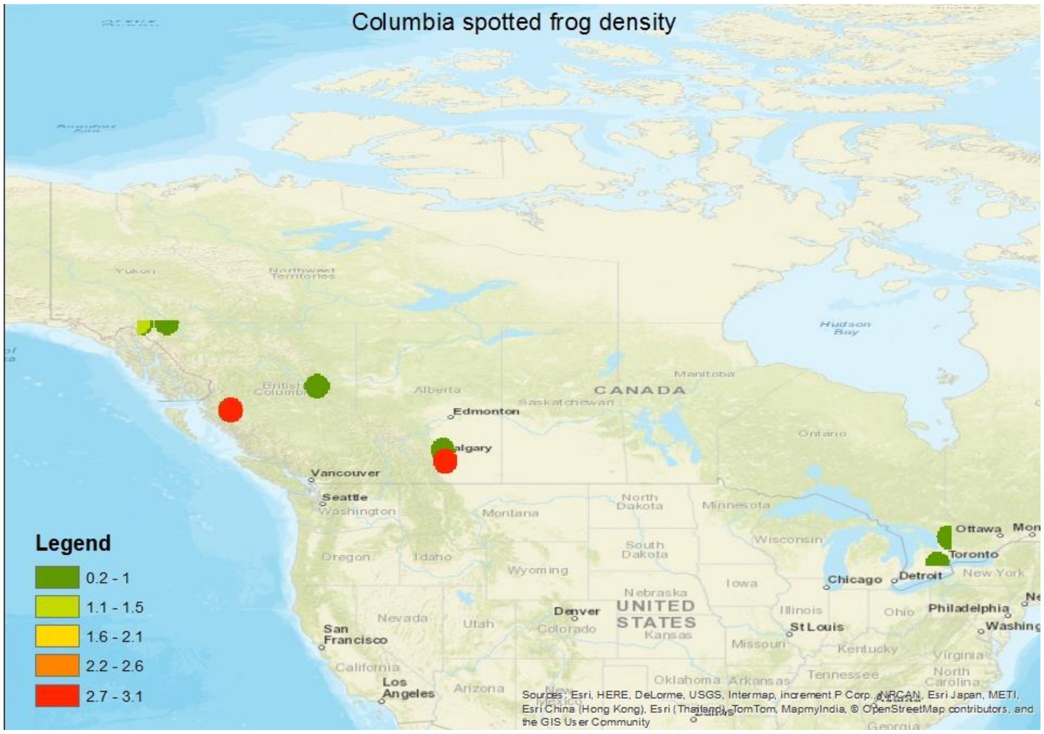

Figure 75: Columbia spotted frog entry density map.

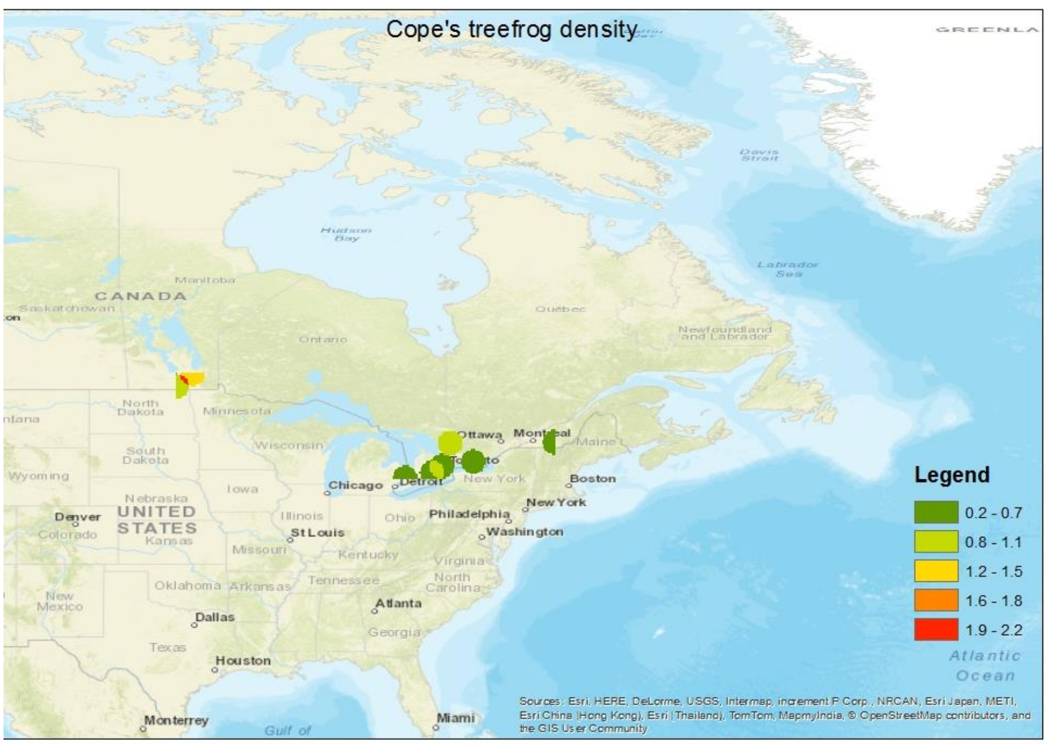

Figure 76: Cope's treefrog entry density map.

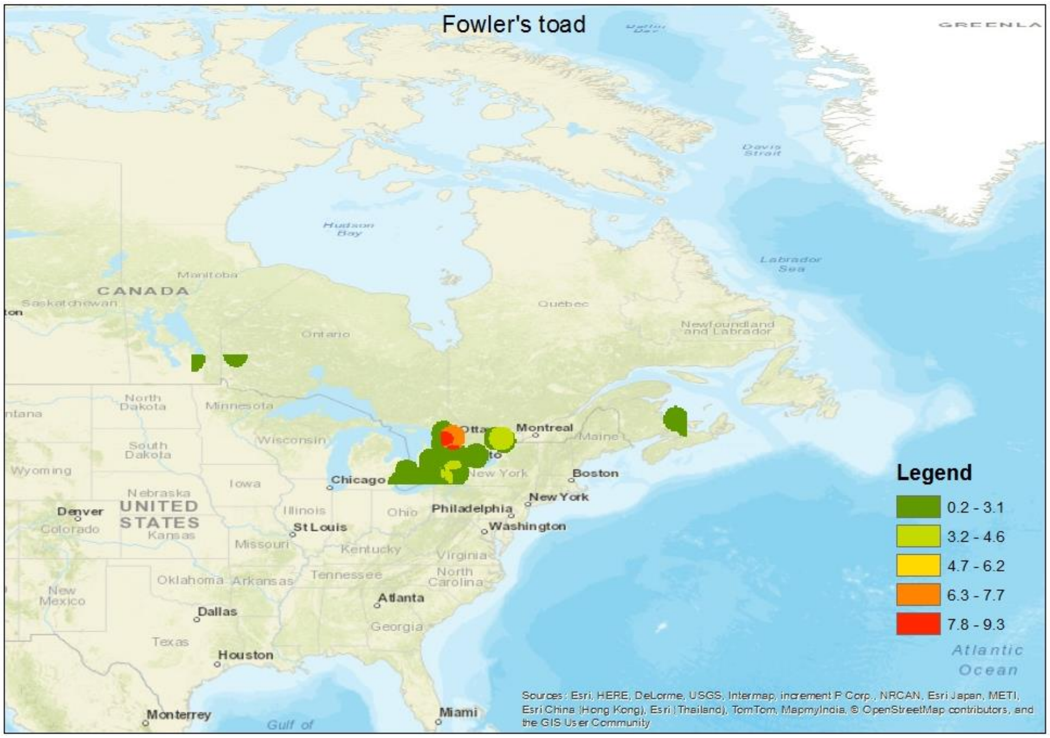

Figure 77: Entry density map for Fowler's toad.

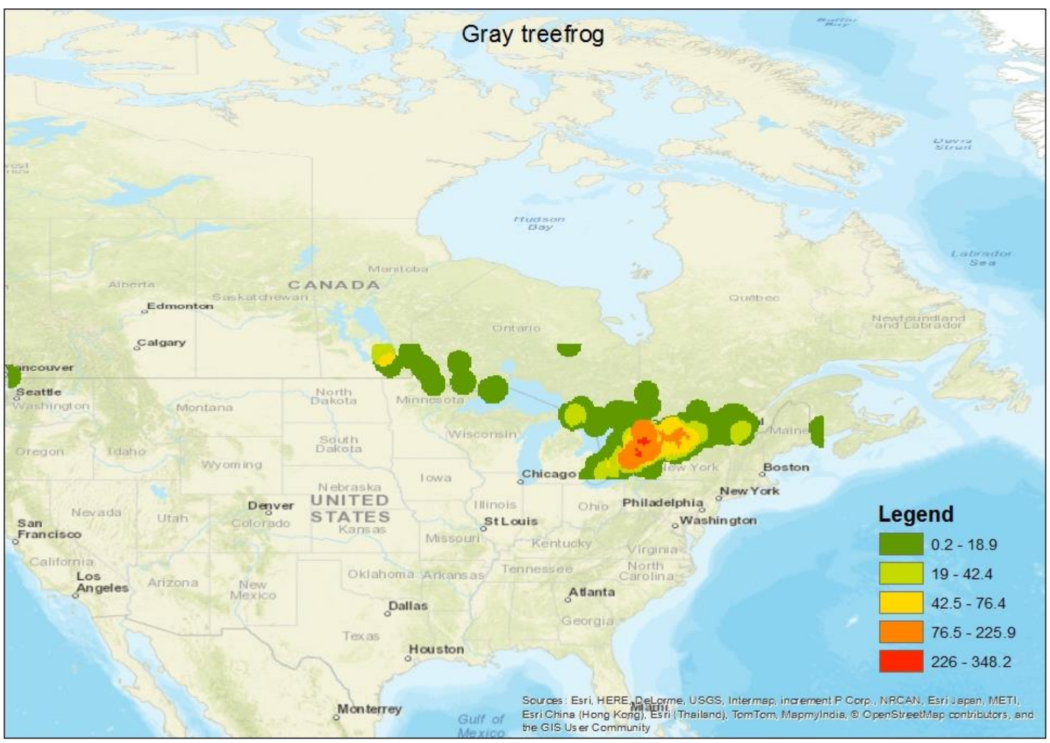

Figure 78: Gray treefrog entry density map.

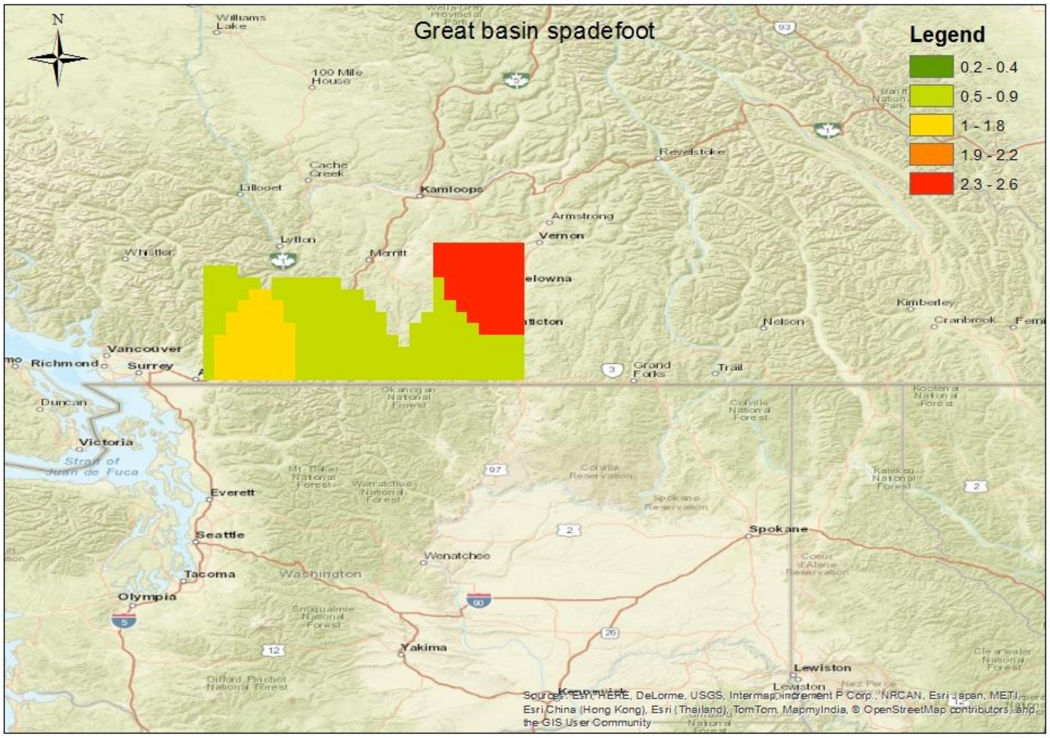

Figure 79: Great Basin spadefoot entry density map.

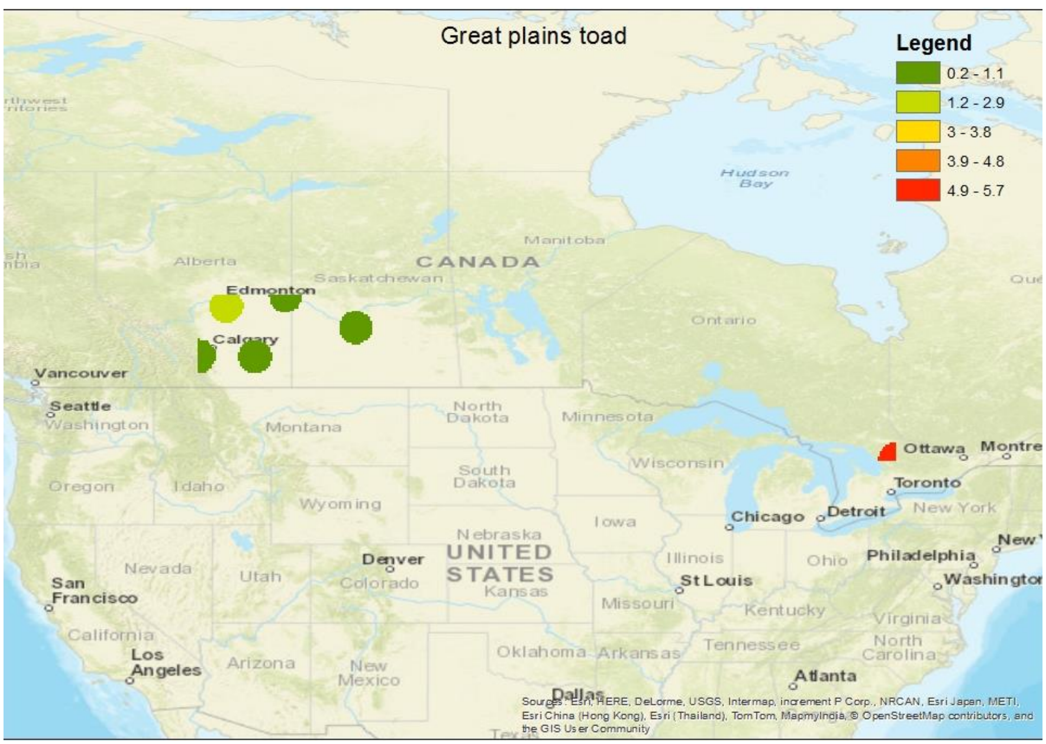

Figure 80: Great plains toad entry density map.

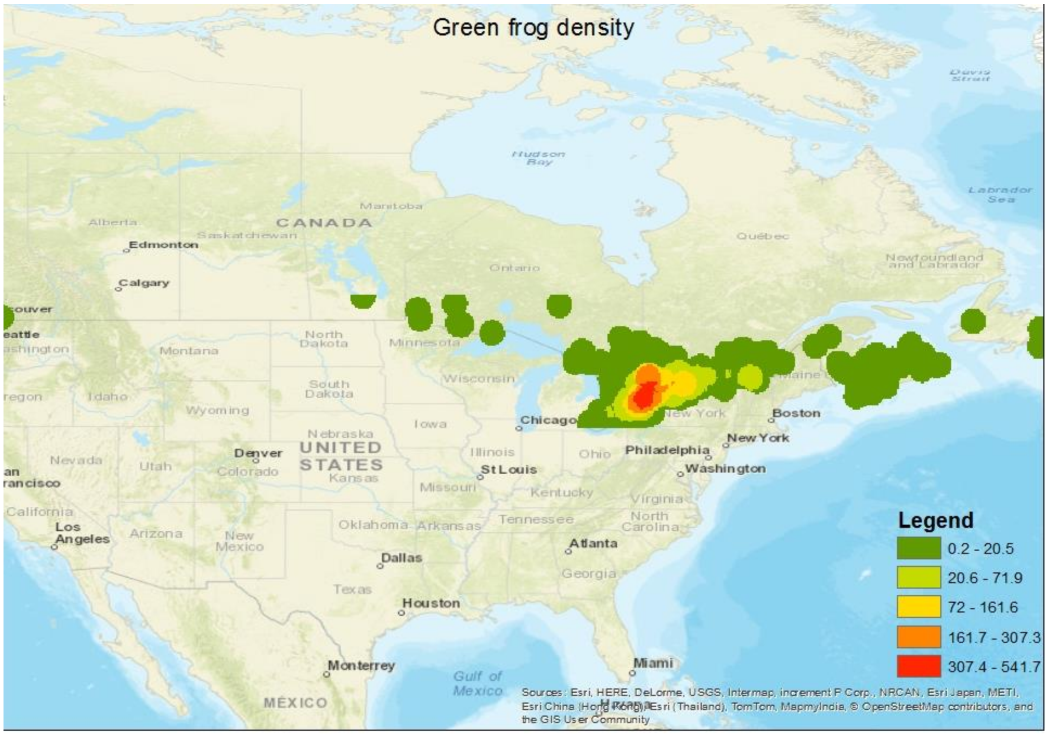

Figure 81: Entry density map for the Green frog.

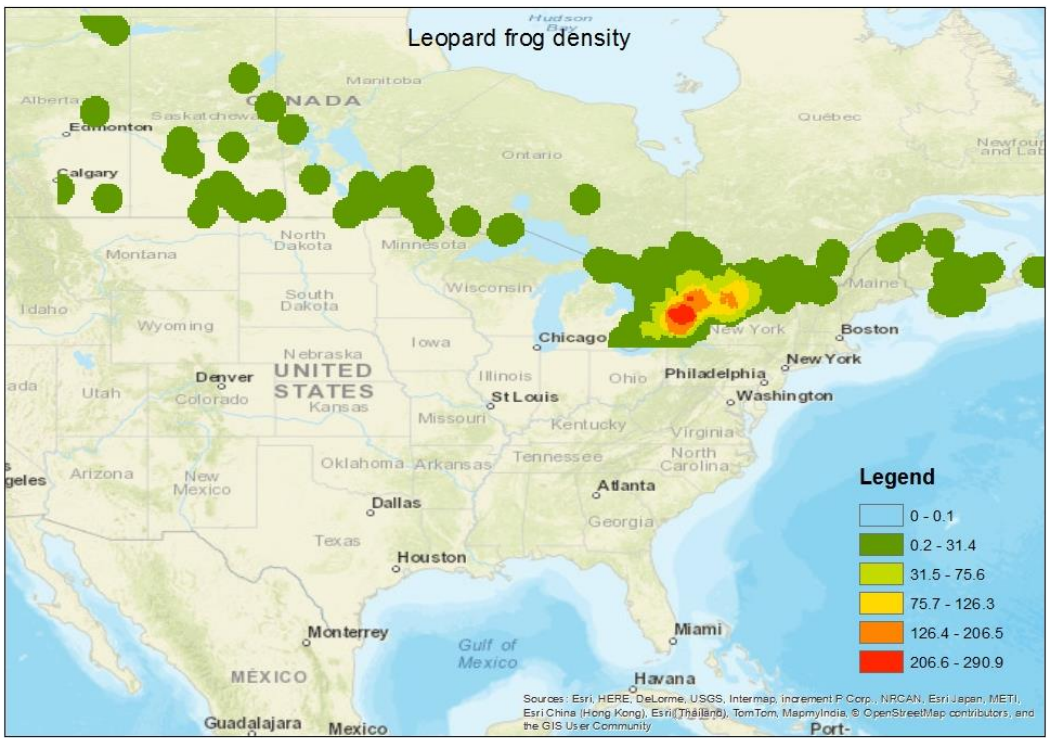

Figure 82: Leopard frog entry density map.

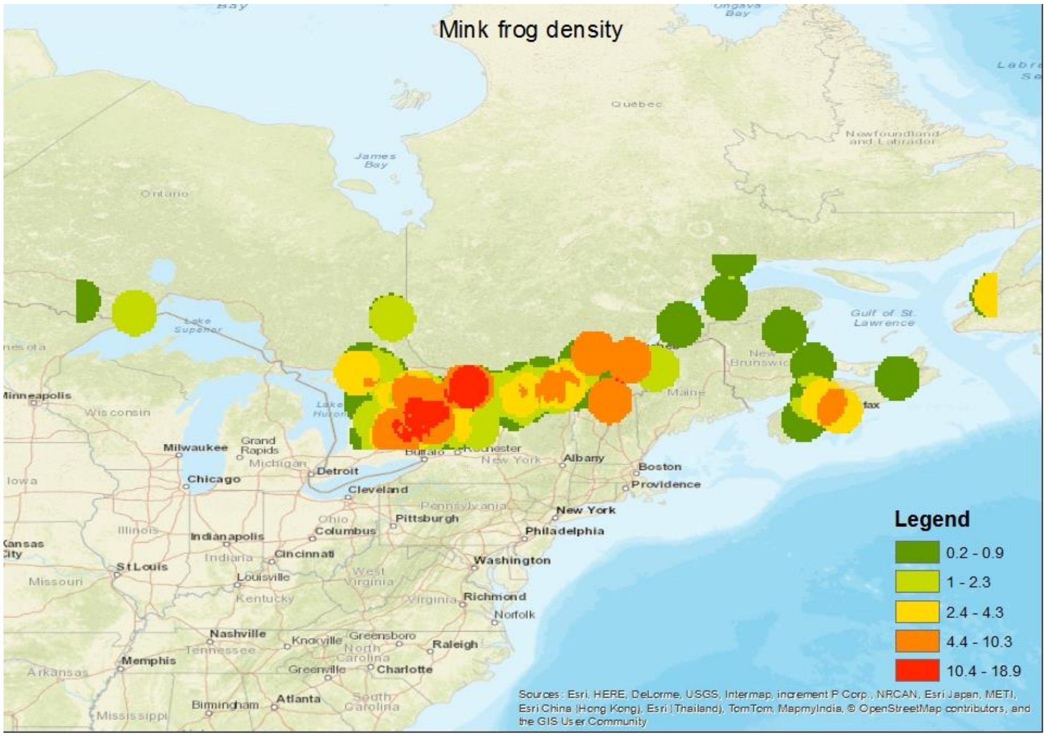

Figure 83: Mink frog entry density map.

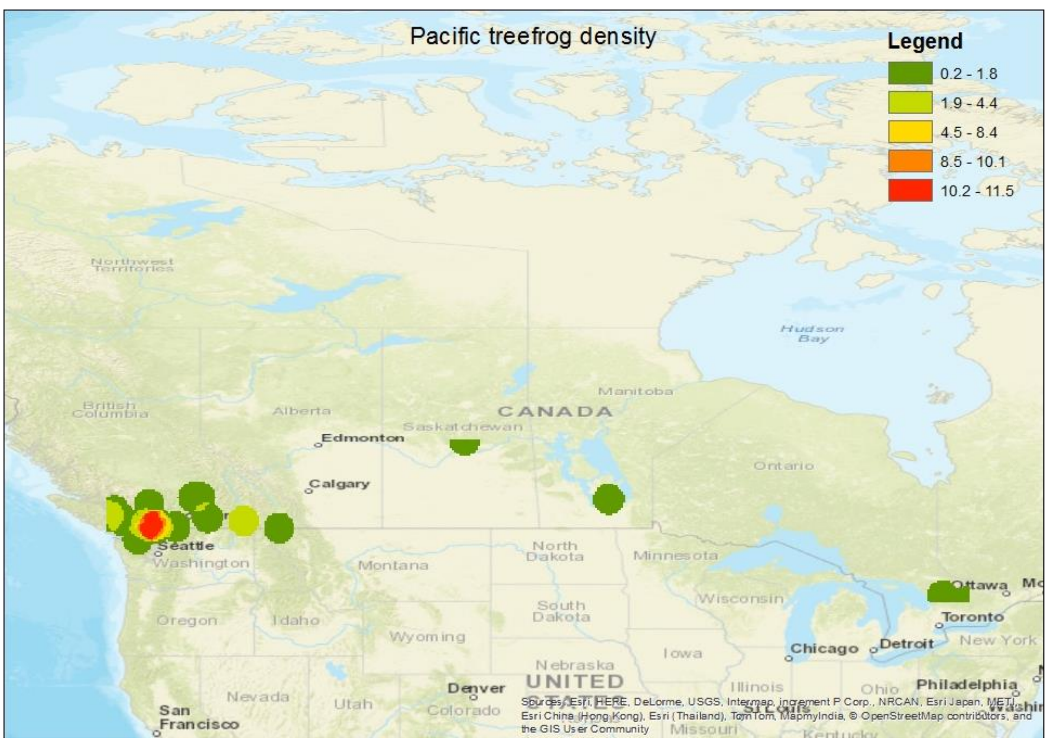

Figure 84: Entry density map for the Pacific treefrog.

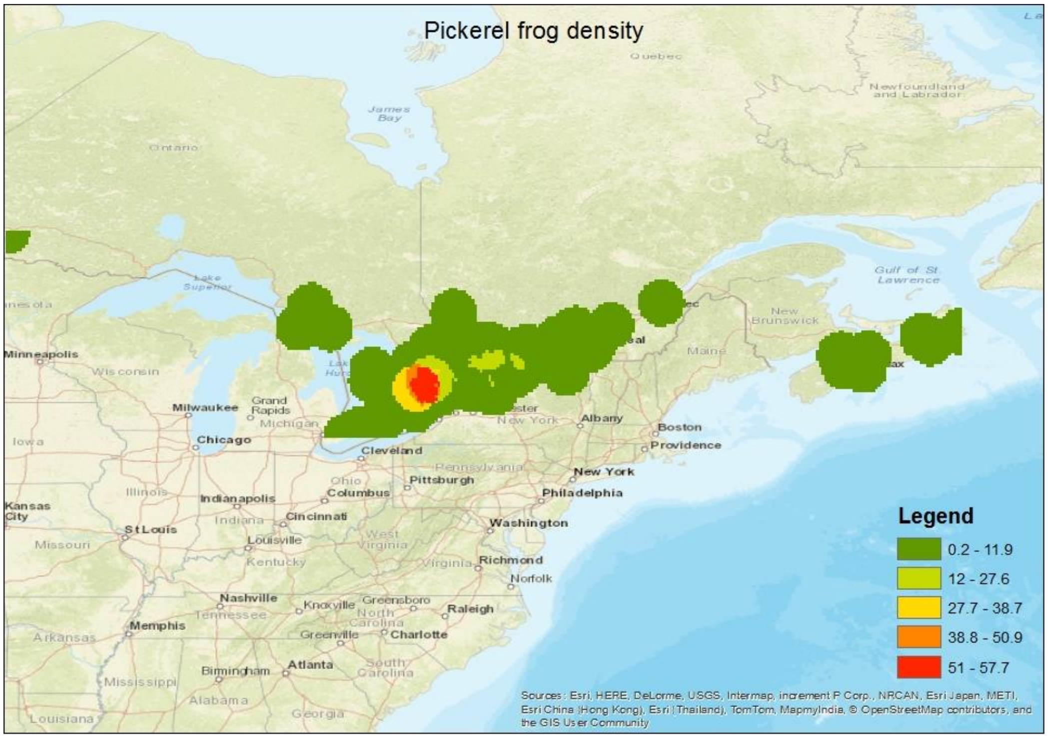

Figure 85: Pickerel frog entry density map.

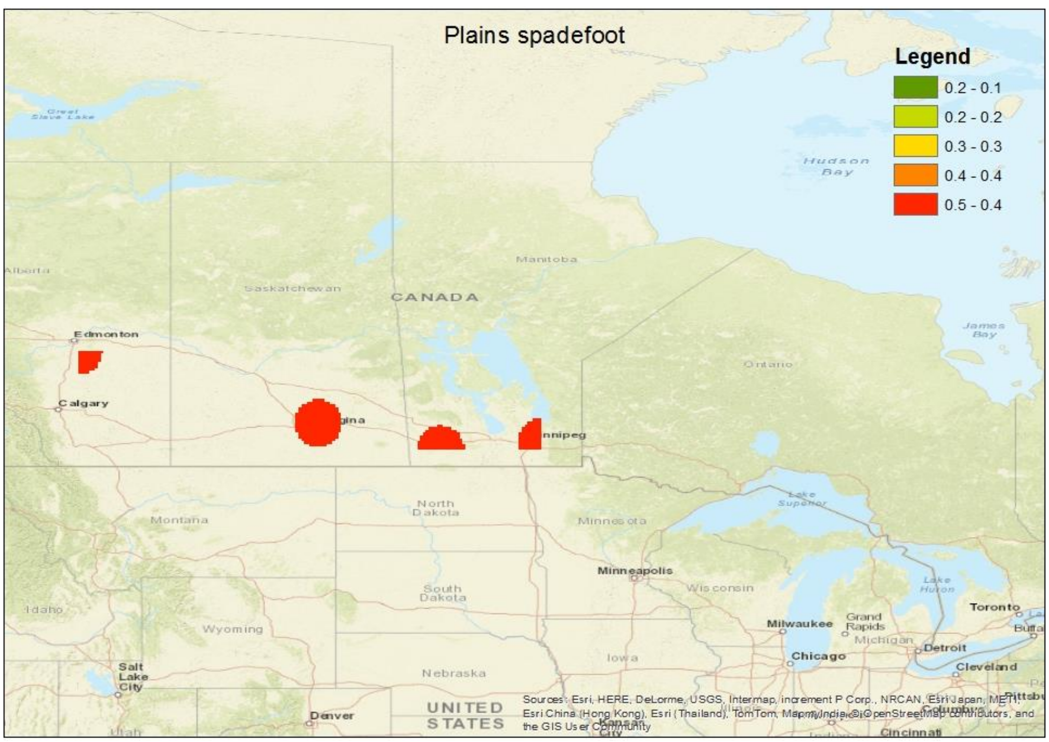

Figure 86: Entry density map for the Plains spadefoot.

Figure 87: Red-legged frog entry density map.

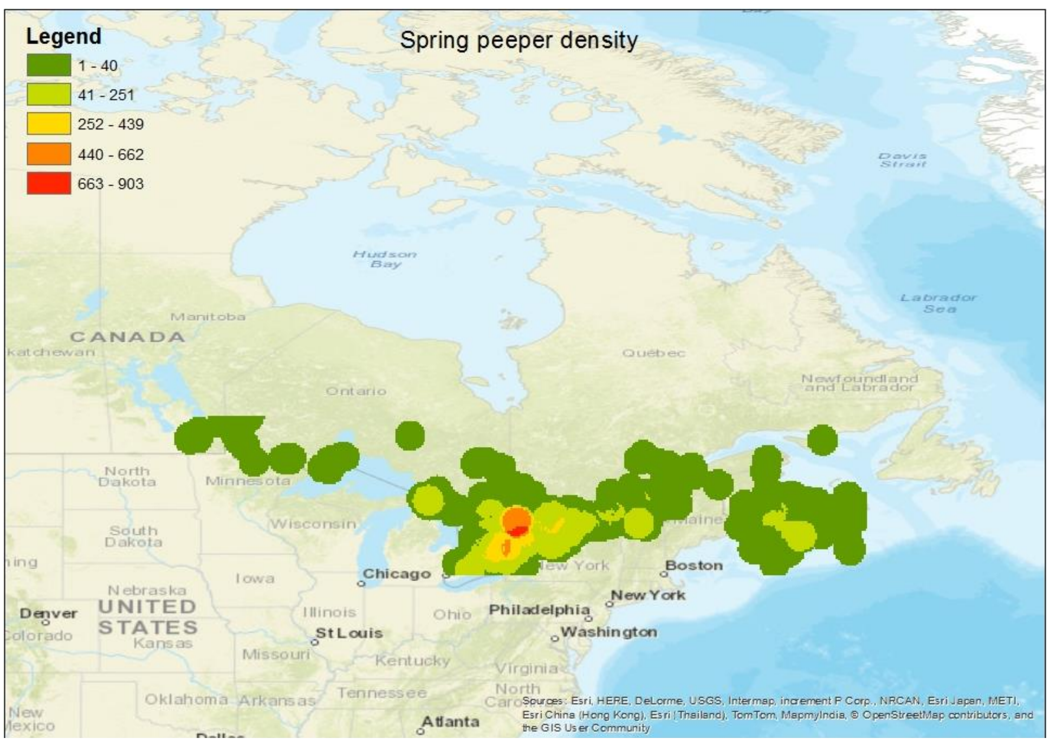

Figure 88: Spring peeper entry density map.

Figure 89: Entry density map for the Tailed frog.

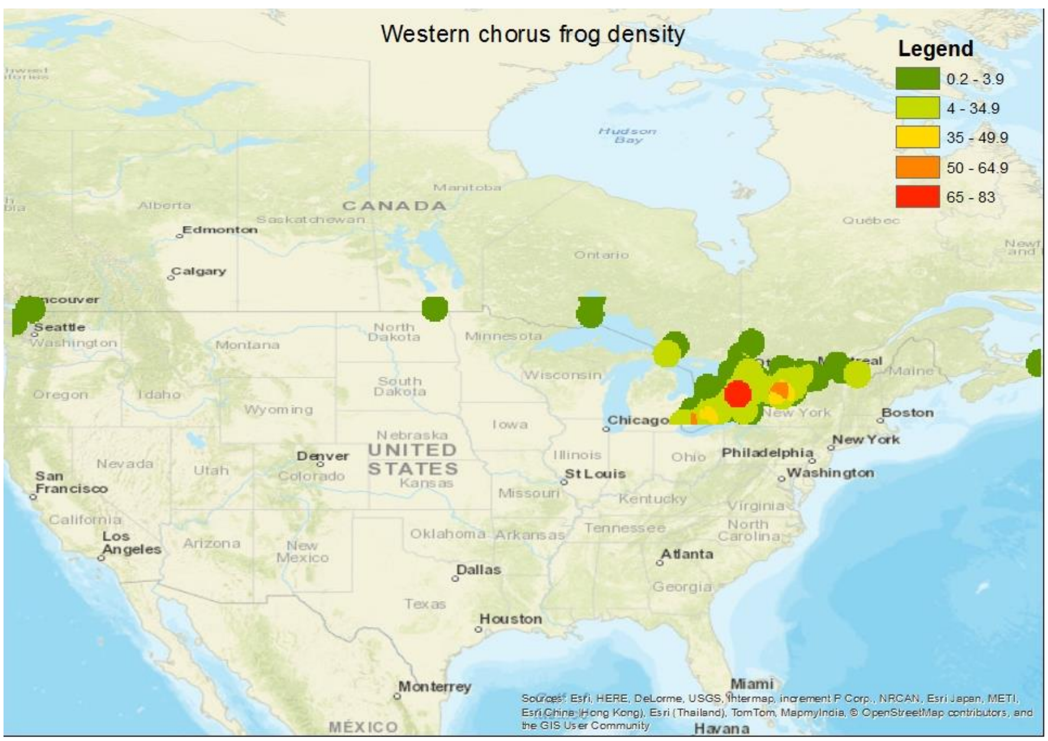

Figure 90: Western chorus frog entry density map.

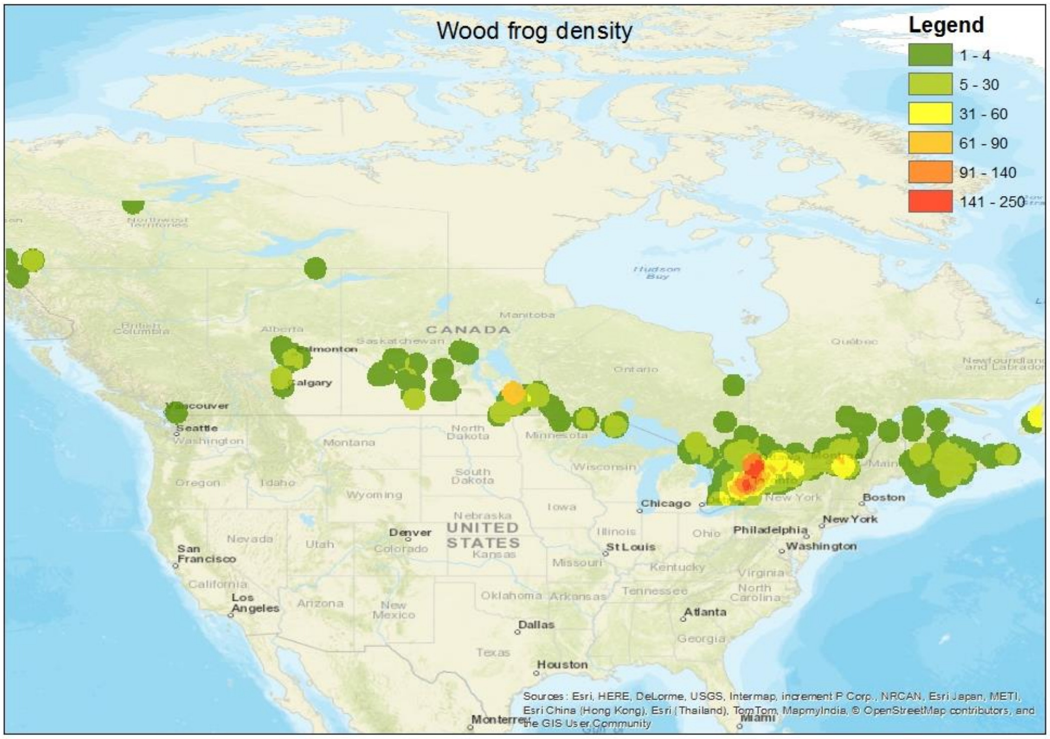

Figure 91: Wood frog entry density map.

The next stage in the analysis is to analyse the observer data, particularly how the entries were spread over the different observers. It is obvious that not every observer will have contributed the same number of entries. This is of interest because identifying where the majority of entries come from is key to deciding how the program is to be promoted. In essence, it helps decide whether appealing more to casual contributors (those with fewer entries) is more productive than targeting dedicated long-term contributors (those with a great many entries). How these vary across the different modules of NatureWatch will also be interesting to see.

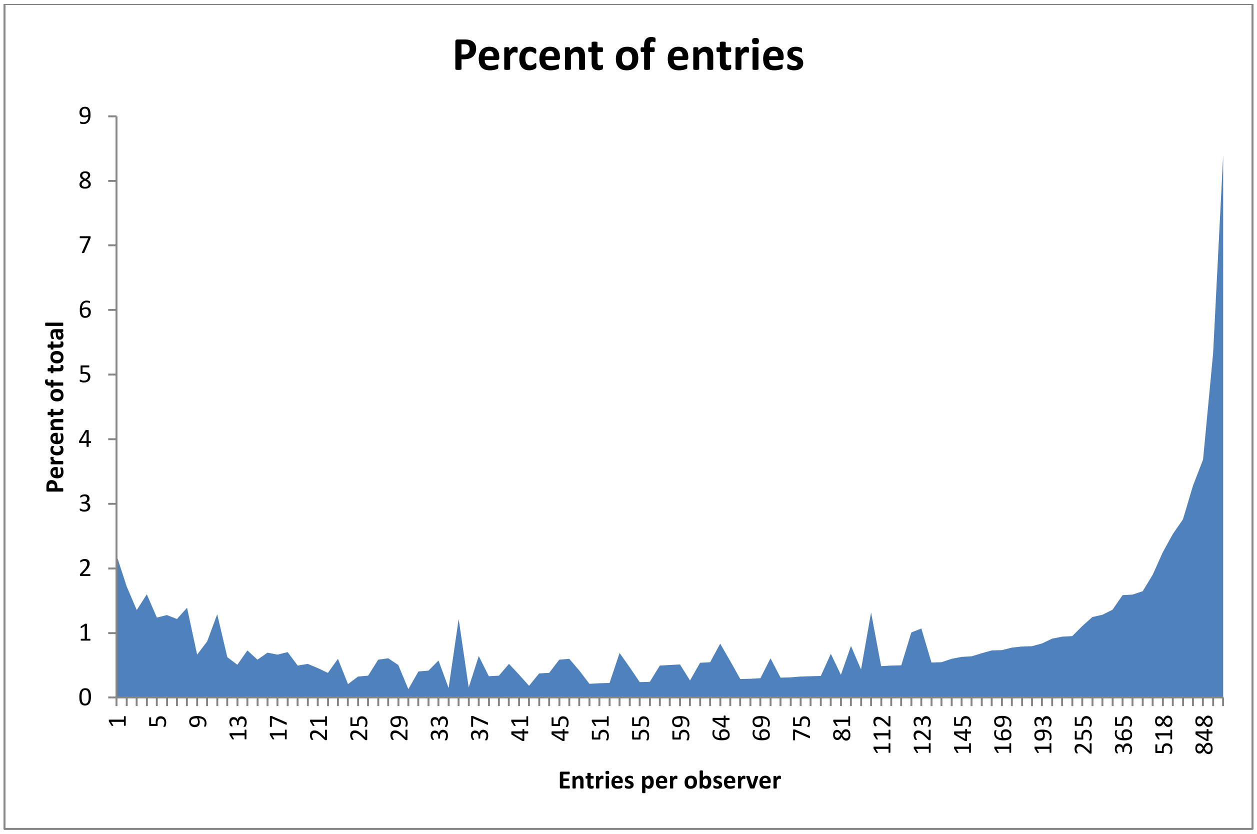

Figure 92: Graph of the percentage of total observations made for each count contribution.

As we can see from the graph above, there seems to be a substantial majority of entries stemming from observers with high entry contributions. In fact, a quick sum of the entry percentages allows us to determine that roughly 13,5% of entries were from contributors with 10 entries or less, whereas those with more than 10 contributed roughly 86,5%. As can be gleaned from the graph as well, these numbers are inflated particularly by those at the extreme high-end, such as the observer who single-handedly contributed 1932 entries (so 8,6% of the total). Unsurprisingly when we consider the percentage of observers who belong in these same contribution brackets, we see that over 80% of observers contributed 10 or fewer entries, and that conversely a little under 20% contributed 10 or more entries. These results quite clearly indicate that most of the entries come from very dedicated observers. While attracting greater numbers of casual observers is desirable as well, it is clear that reaching out to the most prolific observers takes priority for this module.

Appendix B: Plantwatch analysis

Appendix B-1: Plantwatch tables

1.1 plantwatch_observation

Table: plantwatch_observation		
Column heading	**Data type**	**Reference table**
Pk_plantwatch_observation	Integer	
Fk_location	Integer	
Fk_observer	Integer	
Fk_visit	Integer	
Exposure_to_light	Varchar(255)	
Slope	Varchar(255)	
Slope_direction	Varchar(30)	
Fk_plant_species	Integer	
Fk_plantwatch_observation_code	Integer	
Comments	Varchar(255)	
Dandelion_watch	Varchar(25)	
Create_date	Datetime	
Change_date	Datetime	
Visit_date	Datetime	
image	Varchar(255)	

Pk_plantwatch_observation: Number code assigned to the observation based on when it enters the database.
Exposure_to_light: Describes the light the specimen is exposed to.
Slope, Slope_direction: Indicates in general terms what slope the specimen was observed on (flat land, gentle slope or steep slope) and which direction the slope faced where applicable.
Comments: Personalized comments where provided.
Dandelion_watch: Only 12 entries noted, the rest are “NULL”. Defunct column.
Create/change_date: Refers to when the entry was assigned a pk_plantwatch_observation value and when the details have been edited where relevant.
Visit_date: Refers to the date the observation was made as entered by the contributor.
Image: Image, where provided.

1.2 Location

Table: location		
Column heading	**Data type**	**Reference table**
Pk_location	Integer – auto-increment	
Fk_observer	Integer	Observer
Create_date	datetime	
Change_date	datetime	

Location_name	Varchar(255)	
Observer_location_code	Varchar(255)	
Nearest_named_place	Varchar(255)	
Nnp_direction	Varchar(2)	
Nnp_distance	Integer	
latitude	Decimal(18,12)	
longitude	Decimal(18,12)	
Coordinate_source_code	Varchar(2)	
Coordinate_accuracy_code	Varchar(3)	
Location_description	Varchar(255)	
comments	Varchar(255)	
Location_type_code	Char(1)	
elevation	Varchar(20)	
province	Varchar(50)	

Pk_location: Numerical code assigned to each new location, as described in this chapter's introduction.
Create/change_date: Date at which the location was entered into the database and assigned its code.
Location_name: Name of the location as provided by the observer. This can be a general location such as the nearest city or a very specific location such as "[person]'s backyard".
Observer_location_code: Defunct column – none of the entries have non-null values. This is because observers were formerly associated with a single location in the past, while now observers are considered wholly independent from the locations where their observations are made.
Nearest_named_place: This column allows the observer to make locating where the observation was made in the event it is some distance from the "nearest named" place. Usually used in a complimentary manner with the location_name in the event it is too specific/local.
Nnp_direction: Direction to the nnp (nearest named place), using cardinal points.
Nnp_distance: Distance to the nnp.
Latitude/longitude: Coordinates provided by the observer.
Coordinate_source_code: Source of the coordinate data, as provided by the observer. It must be noted that only 191 of the locations have a value entered into this column, out of over 24 500 locations, meaning the coordinates cannot be verified with certainty.
Coordinate_accuracy_code: All values are NULL. Defunct column.
Location_description: Observers can describe the location in this column to provide some additional information.
Comments: The observer can add information here that isn't included in the other columns. All values are NULL.
Location_type_code: 1-digit letter code assigned to the location. Only provided in 100 locations out of the over 24 500, and the data itself is not clearly identified. Defunct column.
Elevation: Where provided, indicates the elevation of the location. Most entries are "NULL".
Province: Name of the province (or territory) where the observation was made.

1.3 Observer

Table: observer		
Column heading	**Data type**	**Reference table**
Pk_observer	Integer – auto-increment	
Fk_partner	Integer	Partner
Fk_age_range	Integer	Age_range
Observer_code	Varchar(255)	

Observer_language	Varchar(1)	
Create_date	Datetime	
Change_date	Datetime	
Observer_first_name	Varchar(255)	
Observer_last_name	Varchar(255)	
Mailing_address_street	Varchar(255)	
Mailing_address_city	Varchar(255)	
Mailing_address_province_code	Varchar(2)	
Mailing_address_country_code	Varchar(3)	
Mailing_address_postal_code	Varchar(255)	
Telephone_number	Varchar(255)	
Email_address	Varchar(128)	
Affiliation	Varchar(128)	
Comments	Varchar(254)	
Send_newsletter	Varchar(10)	
Password	Varchar(512)	
Password_type	Varchar(50)	
Password_temporary	Varchar(16)	
Email_temp	Varchar(255)	
Temp_password_sent	Varchar(255)	
Old_observer_code	Varchar(255)	
Email_validated	Tinyint(4)	

Pk_observer: This column is a numbered identifier for each contributor/observer. The increments are 1 per observer, meaning every observer whose information does not match a pre-existing one is assigned a number 1 higher than the previous one.
Observer_code: Largely analogous to the pk_observer column on a one-for-one basis, except the starting point is much higher (1111111 as opposed to 21).
Observer_language: Possible entries are E and F, for English and French. NULL entries occur when no language is specified.
Create/change_date: Refers to the date the observer was entered into the database, or edited.
Observer_first_name, Observer_last_name: Where provided, identifies the observer's full name via these two columns.
Mailing_address_/city/province_code/country_code/postal_code: Provides the respective mailing address for each category where provided.
Telephone_number, email_address: Where provided, contains the observer's contact information.
Affilitation: The observer's affiliation to an organisation, which can be a private company, corporation or non-profit organisation, among others.
Comments: Personalized comments if any.
Send_newsletter: Notes whether the observer wishes to receive newsletters published on NatureWatch.
Password, Password_type, Password_temporary: The observer's password to sign into the website, for access to the data and to enter observations. The temporary password is the one sent automatically to the observer upon registration, where the "password" variable is null unless the observer changed it to a personalised password.
Email_temp: Temporary password automatically sent to the observer upon registration.

Temp_password_sent: This column informs on whether the observer received a confirmation email for his/her registration as an observer for NatureWatch.
Old_observer_code: The database was altered over the years, including a streamlining of the process of assigning numbers to observers. This column contains the numbers for the observers prior to the update, and so more recently added observers do not have a number in this column.
Email_validated: Denotes if the observer received and acknowledged the confirmation email for the creation of his or her account on the NatureWatch website.

1.4 Visit

Table: visit		
Column heading	**Data type**	**Reference table**
Pk_visit	Integer – auto-increment	
Fk_observation_location	Integer	Observation_location
Fk_owner	Integer	
Fk_watch_programme	Integer	
Create_date	datetime	
Change_date	datetime	
Visit_date	date	
Visit_date_accuracy_code	Varchar(1)	
Air_temperature	Decimal (6,3)	
Water_temperature	Decimal(6,3)	
comments	Varchar(255)	
C_visit_date	Varchar(30)	
Possible_outlier	Varchar(20)	
validated	Varchar(20)	
Audit_date	date	
Fk_administrator_auditor	Integer	

Pk_visit: Number assigned to the visit of a particular patch of plants (differs from Frogwatch because subsequent visits of the same patch of plants or water body are possible and easy to carry out)
Create/change_date: Date where the pk_visit entry was entered or changed (i.e. does not correspond to the actual visit date).
Fk_owner: Defunct column, the foreign key connects nowhere.
Visit_date: Date where the visit was carried out, as entered by the observer.
Visit_date_accuracy_code: Defunct column, only 338 entries out of over 47 000 are not “NULL”.
Air/water_temperature: The air and water temperature entered by the observer, water temperature only where relevant (table is also used in Icewatch).
Comments:
C_visit_date: Serves as a confirmation of the visit date. The purpose of this variable is unclear, as most of the entries have a NULL value.
Possible_outlier: Identifies whether the entry is an outlier. 60 entries out of over 44 000 are considered as such.
Validated: Only 2 entries in this column are labeled “true” in this column, the vast majority being “NULL” with a number of “false” entries. Defunct column.
Audit_date: Defunct column – the only non-“NULL” value being for a “test” entry.
Fk_administrator_auditor: Defunct column – the foreign key leads nowhere.

1.5 plant_species

Table: plant_species		
Column heading	**Data type**	**Reference table**
Pk_plant_species	Integer	
Common_english_name	Varchar(255)	
Common_french_name	Varchar(255)	
Scientific_name	Varchar(255)	

Pk_plant_species: Number code used for a given species in the database.
Common_english/French/scientific_name: The name the code corresponds to in English and French, along with the scientific name of the species.

1.6 plantwatch_observation_code

Table: plantwatch_observation_code		
Column heading	**Data type**	**Reference table**
Pk_plantwatch_observation_code	Integer	
Eng_description	Varchar(255)	
Fre_description	Varchar(255)	

Pk_plantwatch_observation_code: A code assigned to the circumstances of the observation, being the number and surroundings of the plant(s).
Eng/fre_description: English and French descriptions of the circumstances of the observation.

1.7 partner

Table: partner		
Column heading	**Data type**	**Reference table**
Pk_partner	Integer	
Create_date	Datetime	
Change_date	Datetime	
English_name	Varchar(255)	
French_name	Varchar(255)	
Mailing_address_street	Varchar(255)	
Mailing_address_city	Varchar(64)	
Mailing_address_province_code	Varchar(2)	
Mailing_address_country_code	Varchar(3)	
Mailing_address_postal_code	Varchar(6)	
Telephone_number	Varchar(10)	
Fax_number	Varchar(10)	
Email_address	Varchar(128)	
Contact_name	Varchar(128)	
Comments	Varchar(255)	

Pk_partner: Number code assigned to the program partner.
Create/change_date: Date when the pk_partner entry was entered or its details changed.
English/French_name: English and French names of the partner organisation.
Mailing_address_street/city/province_code/country_code/postal_code: Coordinates of the partner.

Telephone/fax_number, email_address: Contact information.
Comments: Short description or comment, where relevant.

It must be noted that there are only 4 entries in this table, none of which contain any information besides the pk_partner and create/change date. The table as a whole is thus defunct.

1.8 age_range

Table: age_range		
Column heading	**Data type**	**Reference table**
Pk_age_range	Integer	
Range_text	Varchar(50)	

Pk_age_range: Assigns an increment code to the age of the observer.
Range_text: The values (high and low) delimiting the above increments.

1.9 watch_programme

Table: watch_programme		
Column heading	**Data type**	**Reference table**
Pk_watch_programme	Integer	
Create_date	Date	
Change_date	Date	
english_name	Varchar(255)	
french_name	Varchar(255)	
English_description	Varchar(255)	
French_description	Varchar(255)	

Pk_watch_programme: Number code assigned to the programme for which an observation was entered.
Create/change_date: Date where the program's pk_watch_programme value was entered, and for when its information was edited.
English/French_name: Name of the program in the specified language.
English/French_description: A brief description of the program in the specified language, where provided.

Appendix B -2: Plantwatch trend analysis

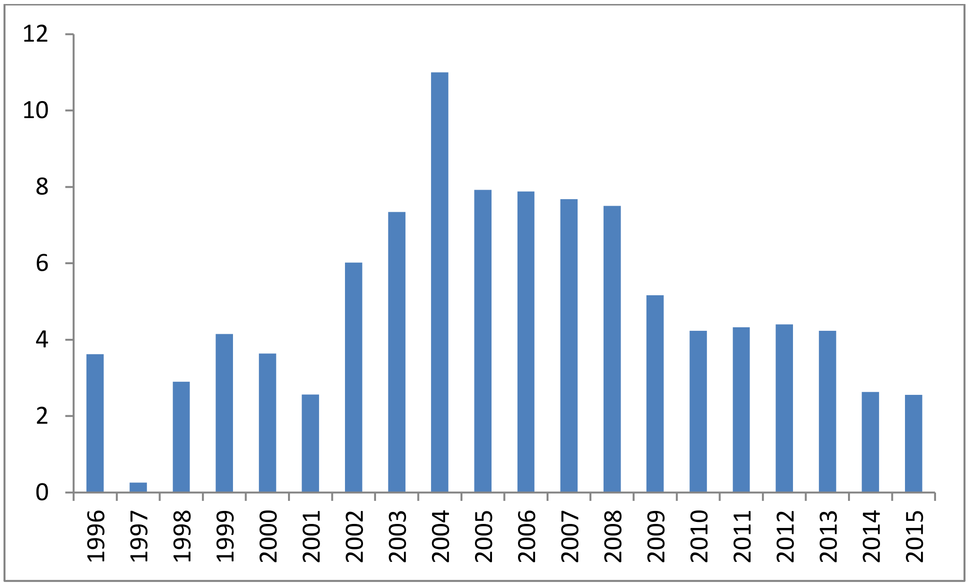

Figure 93: Frogwatch aggregate entry counts by year.

As we can see from this graph, there have been declining plateaux since 2008. While the top year was 2004, there were clearly "good years" from 2002 to 2008 for this program.

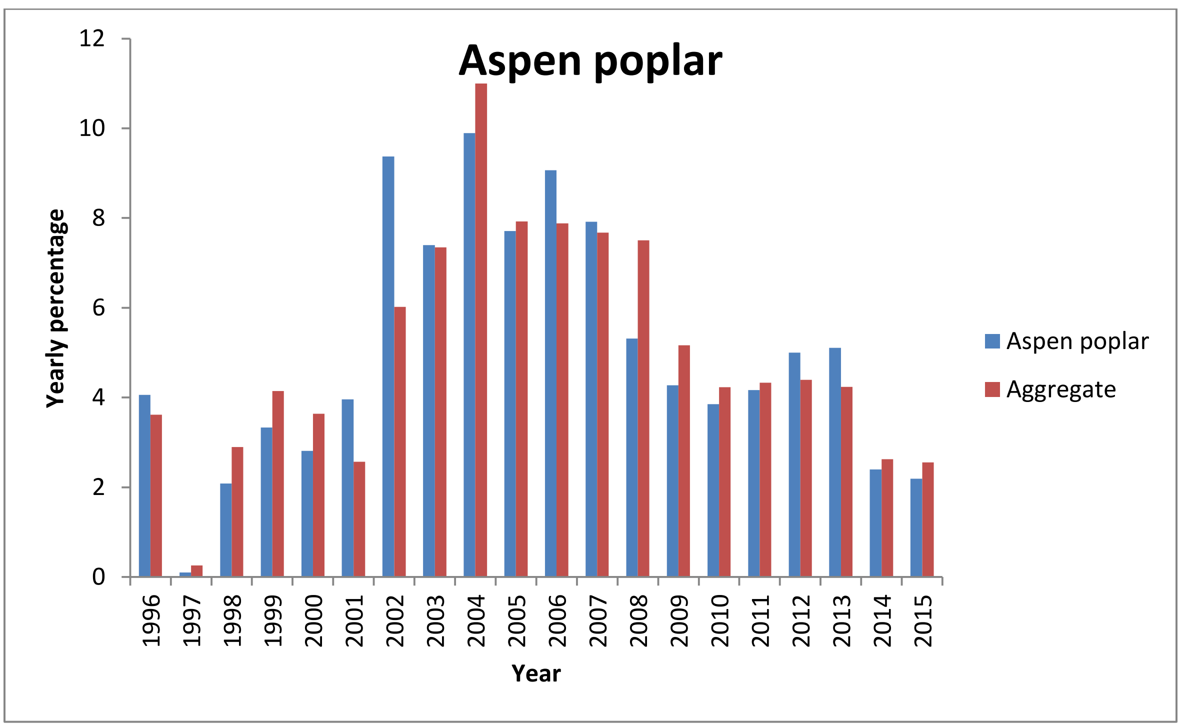

Figure 94: Yearly percentage of the total number of entries for the aggregate and those for Aspen poplar.

The number of entries for this species, being 983, means it has fairly good value for analyses. As we can see it mostly follows the aggregate percentage values, except for 2002 where it rises above the aggregate. This species may be of interest for analysis as a result of its overall consistency except for this year.

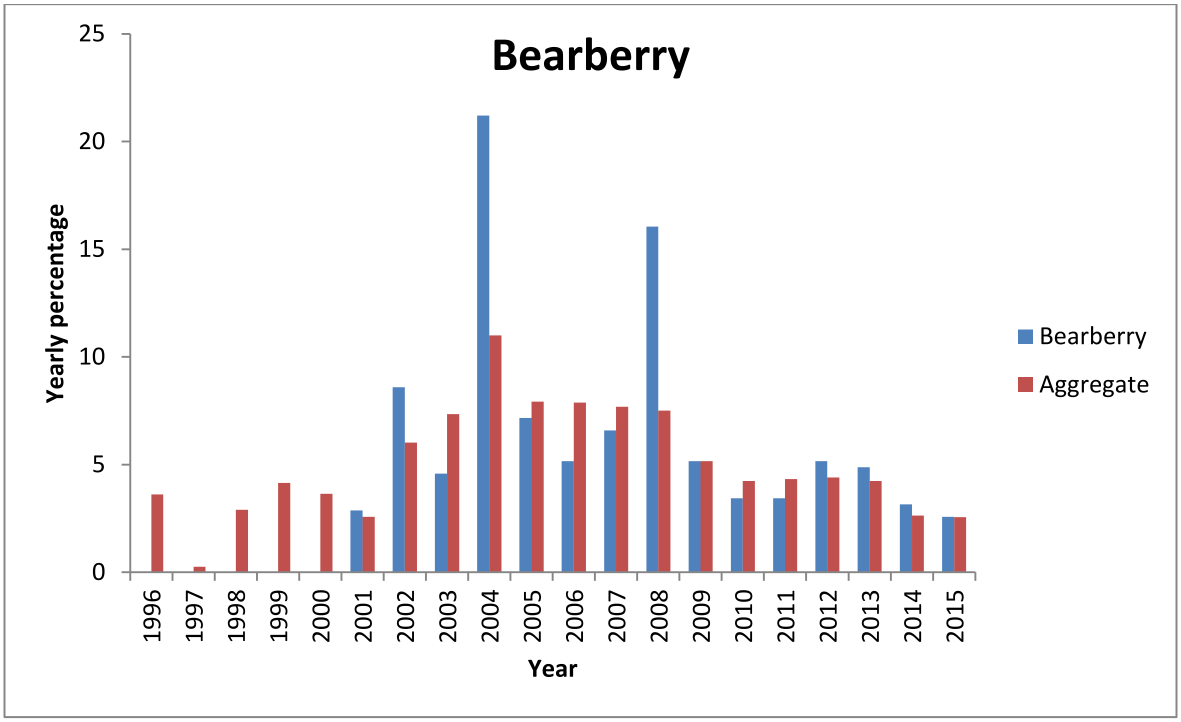

Figure 95: Yearly percentage of the total number of entries for the aggregate and those for Bearberry.

At 349 total entries before 2016, this species is not ideal for analysis. Most of its entries occurred in 2004 and 2008.

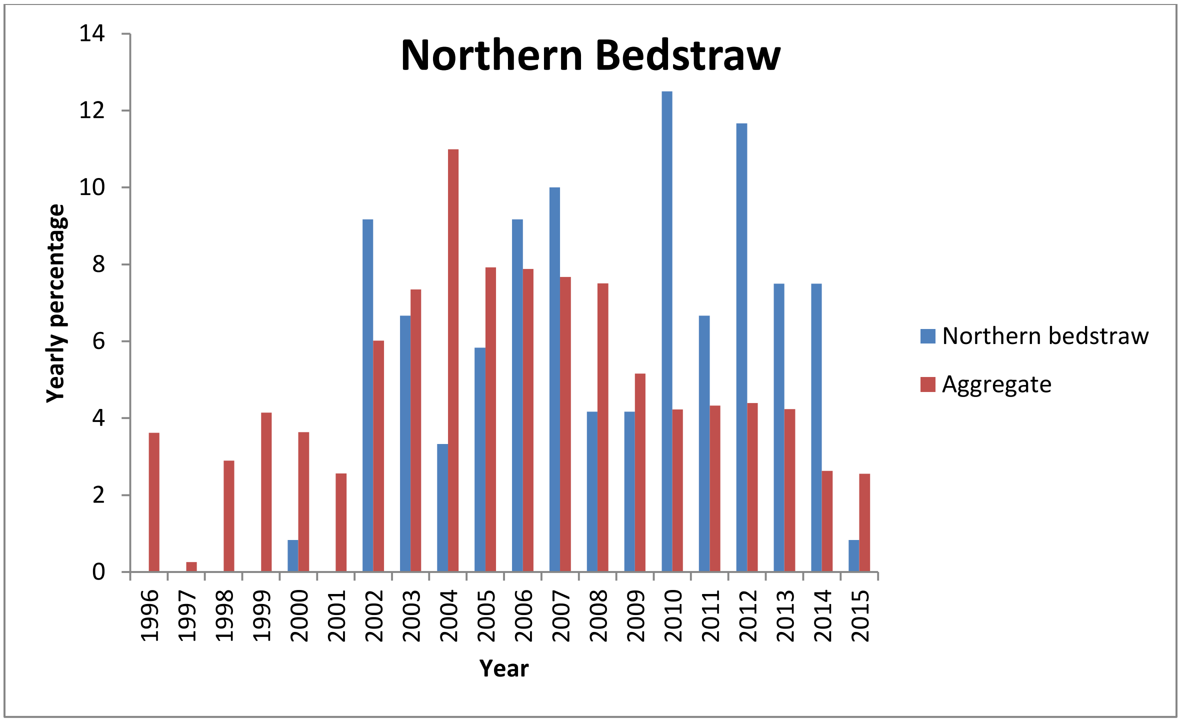

Figure 96: Yearly percentage of the total number of entries for the aggregate and those for Northern bedstraw.

With only 121 entries before 2016, this species has too few entries to be of much use in analyses.

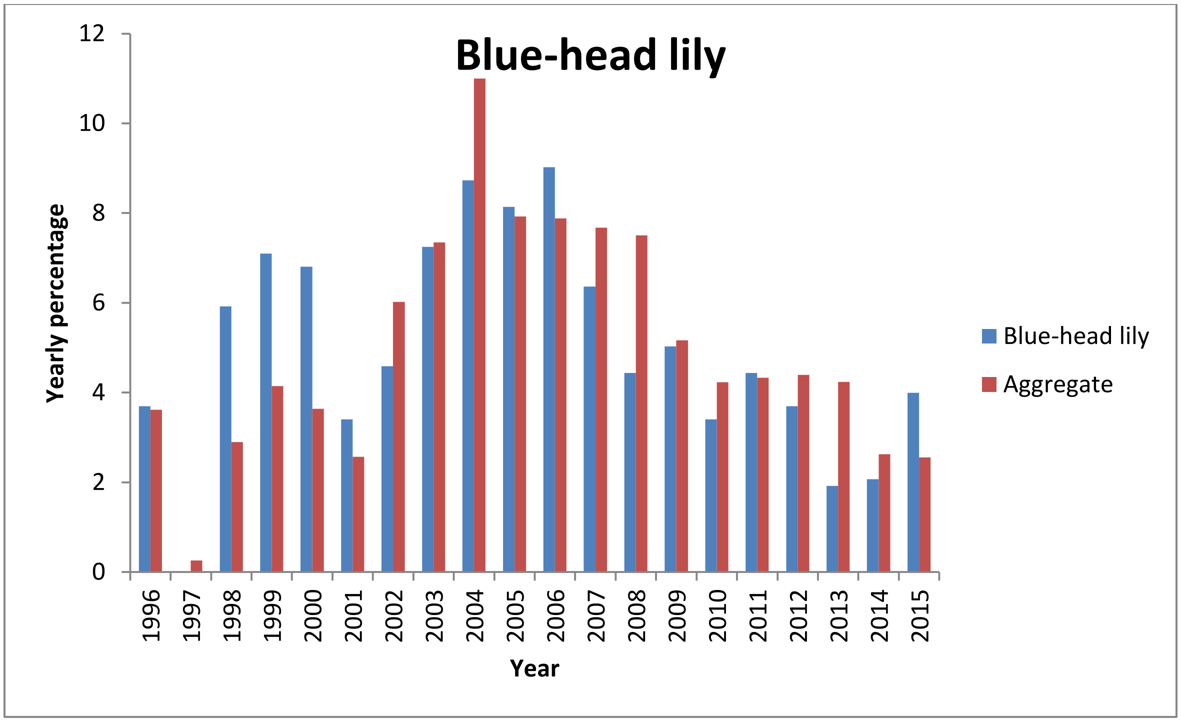

Figure 97: Yearly percentage of the total number of entries for the aggregate and those for Blue-head lily.

This species has some 681 entries before 2016, as such it is reasonably valuable for analyses. It has high values from 1998 to 2000.

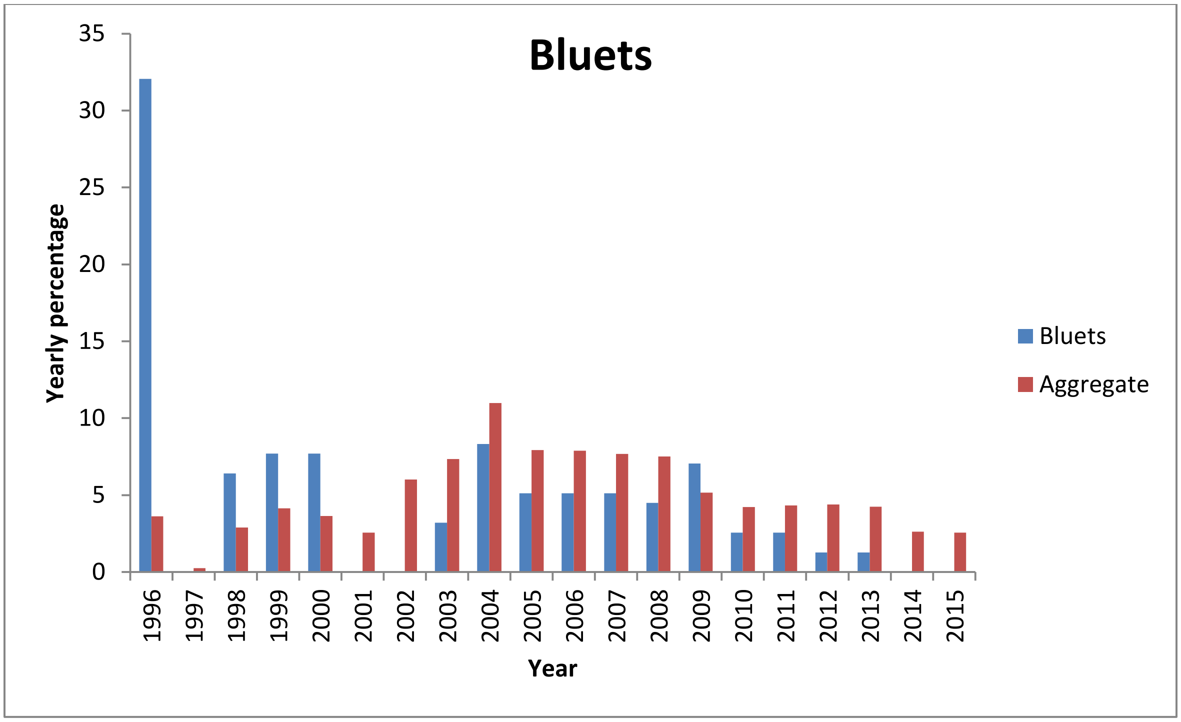

Figure 98: Yearly percentage of the total number of entries for the aggregate and those for bluets.

The bluets, with 163 entries only, are not very good for analyses due to the small samples. A significant portion of the entries are in 1996 alone.

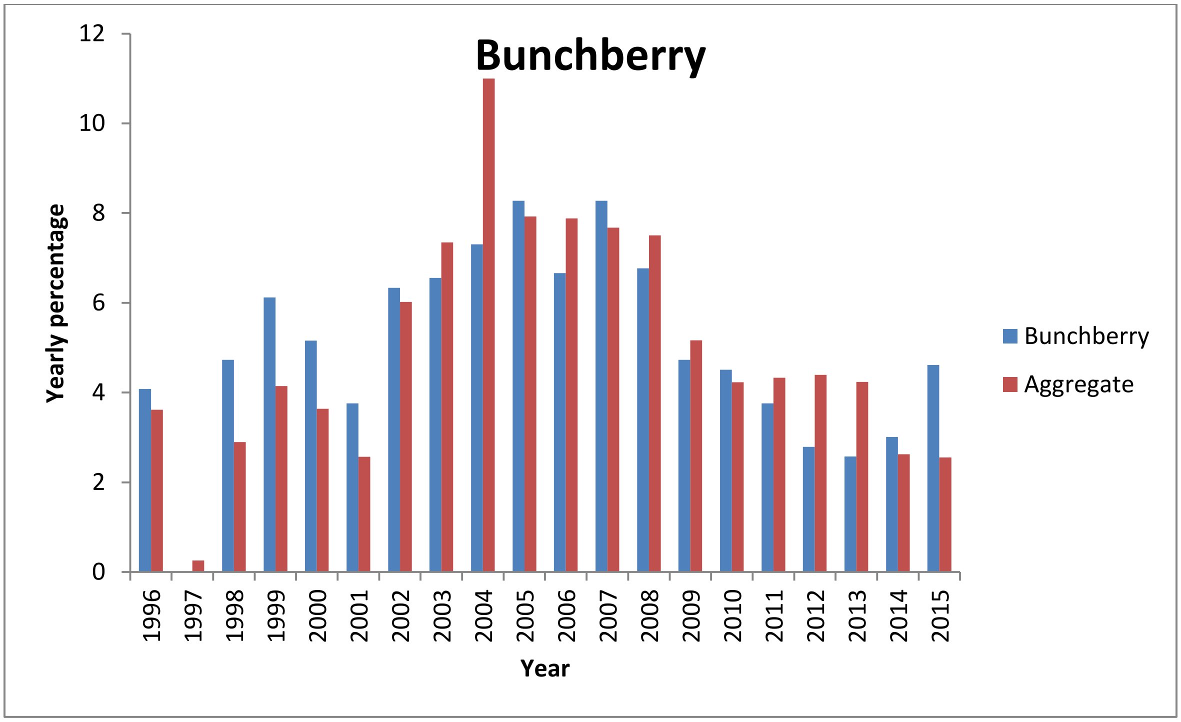

Figure 99: Yearly percentage of the total number of entries for the aggregate and those for Bunchberry.

This species has an appreciable entry count with 953 before 2016. Interestingly it does not follow the 2004 spike in the aggregate.

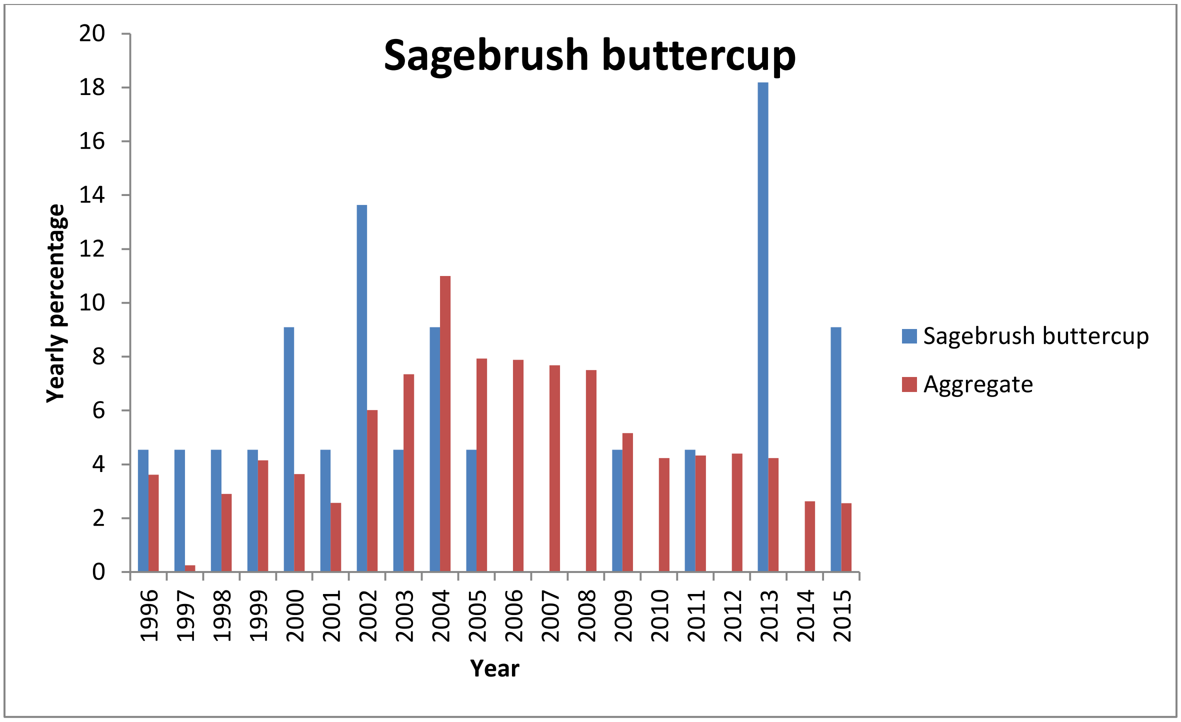

Figure 100: Yearly percentage of the total number of entries for the aggregate and those for Sagebrush buttercup.

This species has only 26 entries.

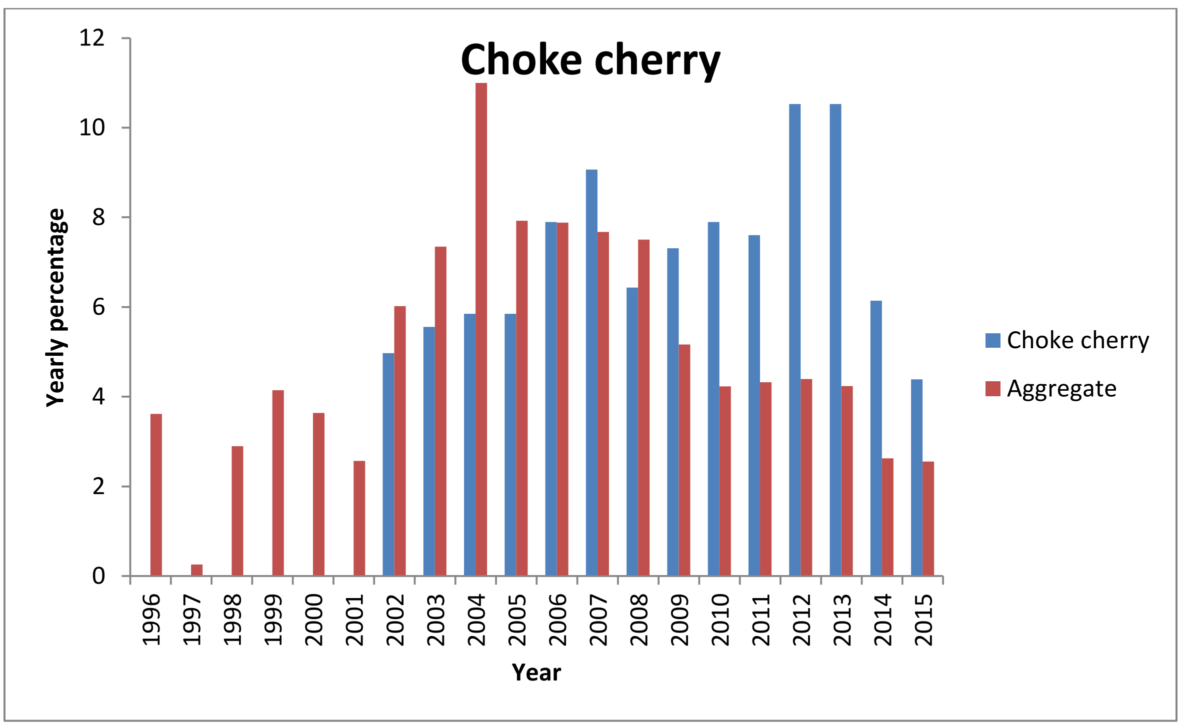

Figure 101: Yearly percentage of the total number of entries for the aggregate and those for Choke cherry.

The Choke cherry has some 346 entries, which means it is not very useful for analysis. It also diverges notably from the aggregate on most years, with high values in recent years.

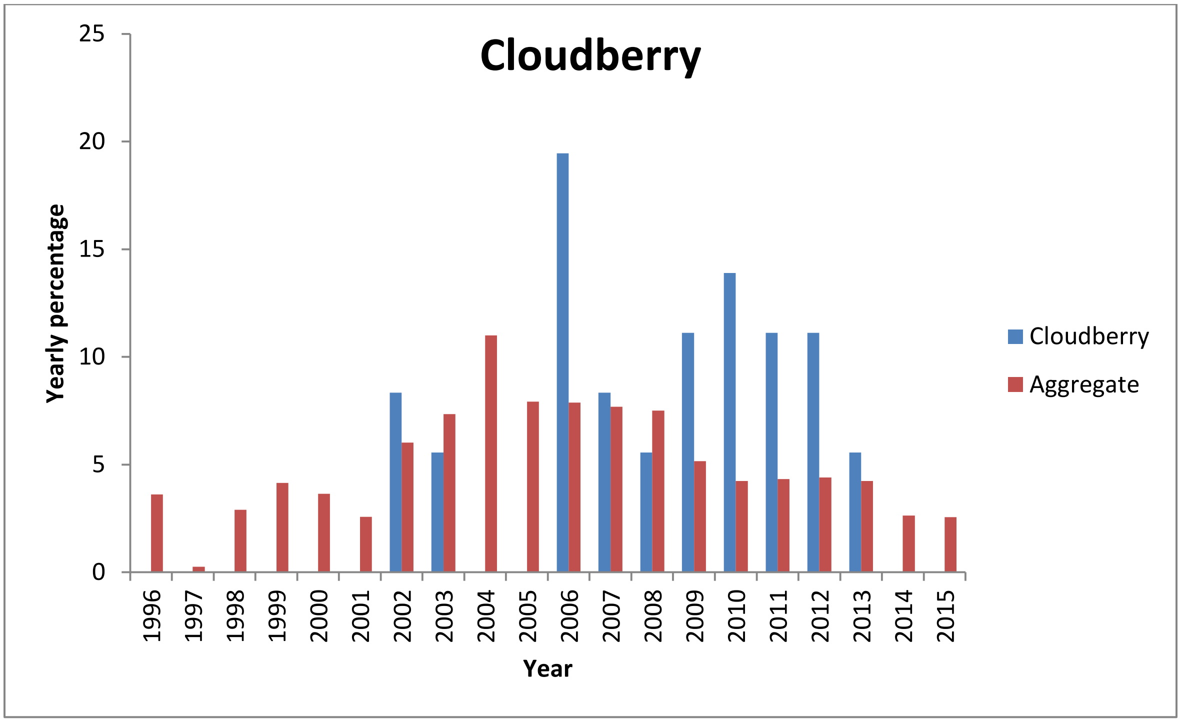

Figure 102: Yearly percentage of the total number of entries for the aggregate and those for Cloudberry.

This species has had 36 entries before 2016, which is not enough, especially considering the number of years where there were no entries at all.

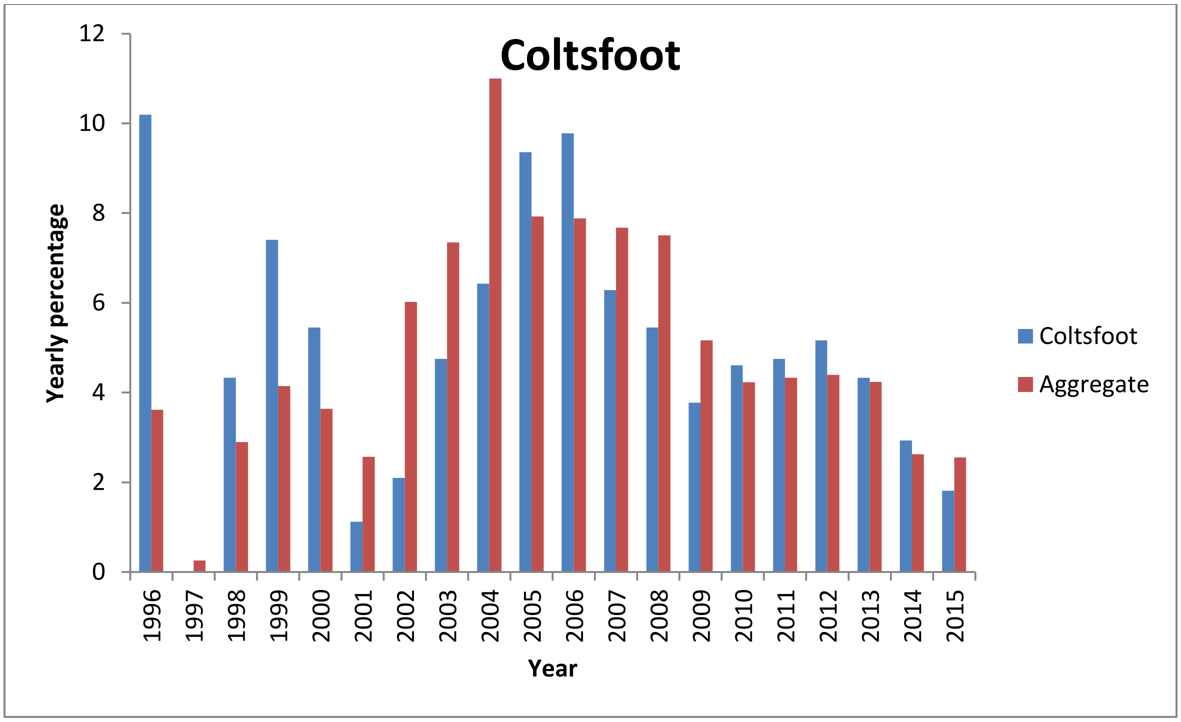

Figure 103: Yearly percentage of the total number of entries for the aggregate and those for Coltsfoot.

The 726 entries is actually good for analysis. Interestingly this species does not reflect the 2009 aggregate spike at all. It spikes in 2010 and 2012, and besides these years it closely follows the aggregate.

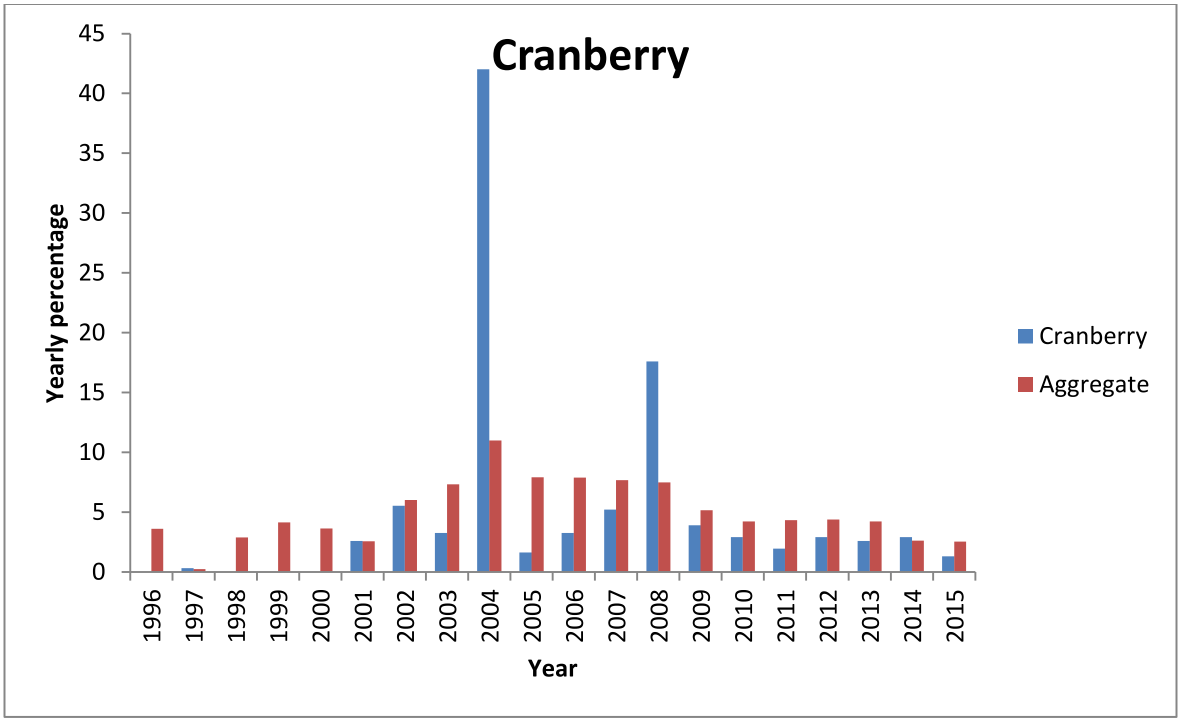

Figure 104: Yearly percentage of the total number of entries for the aggregate and those for Cranberry.

Cranberry has a total of 307 entries prior to 2016. The 2004 spike is too large for it to be considered for analysis.

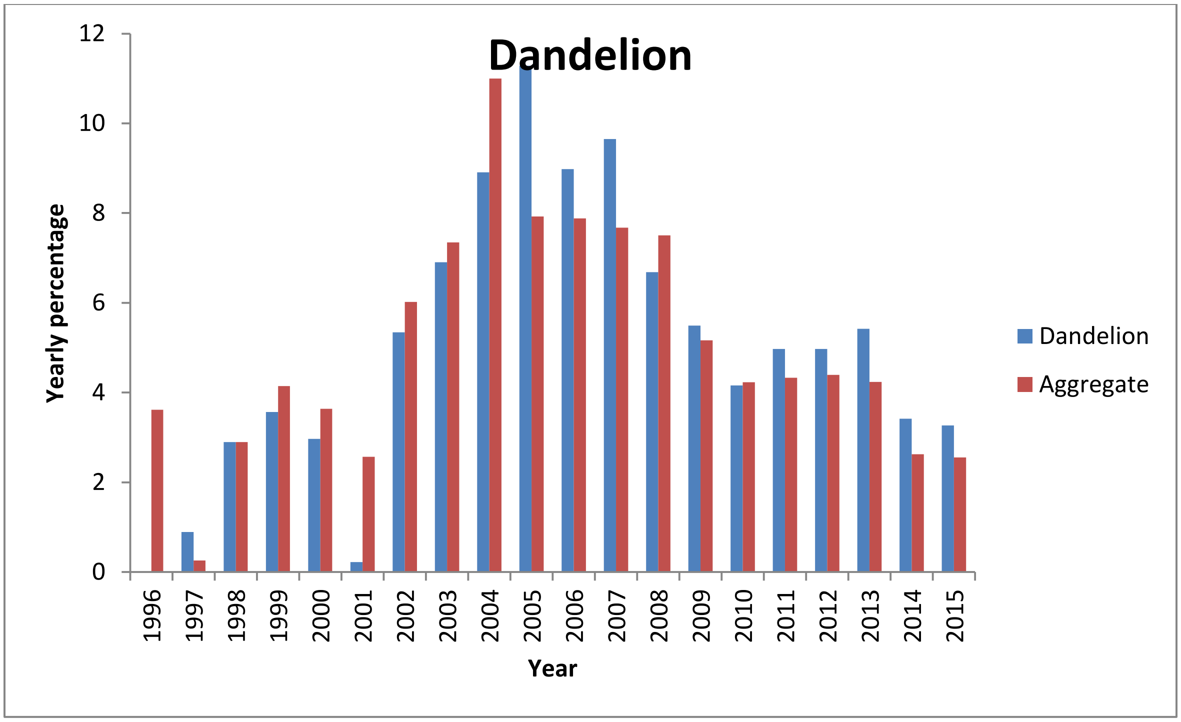

Figure 105: Yearly percentage of the total number of entries for the aggregate and those for Dandelion.

This introduced species is extremely abundant (with over 1366 entries before 2016). It also generally follows the aggregate very closely except for 2004 and 2005. Given how widespread it is and how it thrives in human-made lawns unless managed it is not the best species for analysis though, except perhaps as a means to isolate the entry counts for other species from participation (as it is so easy to find).

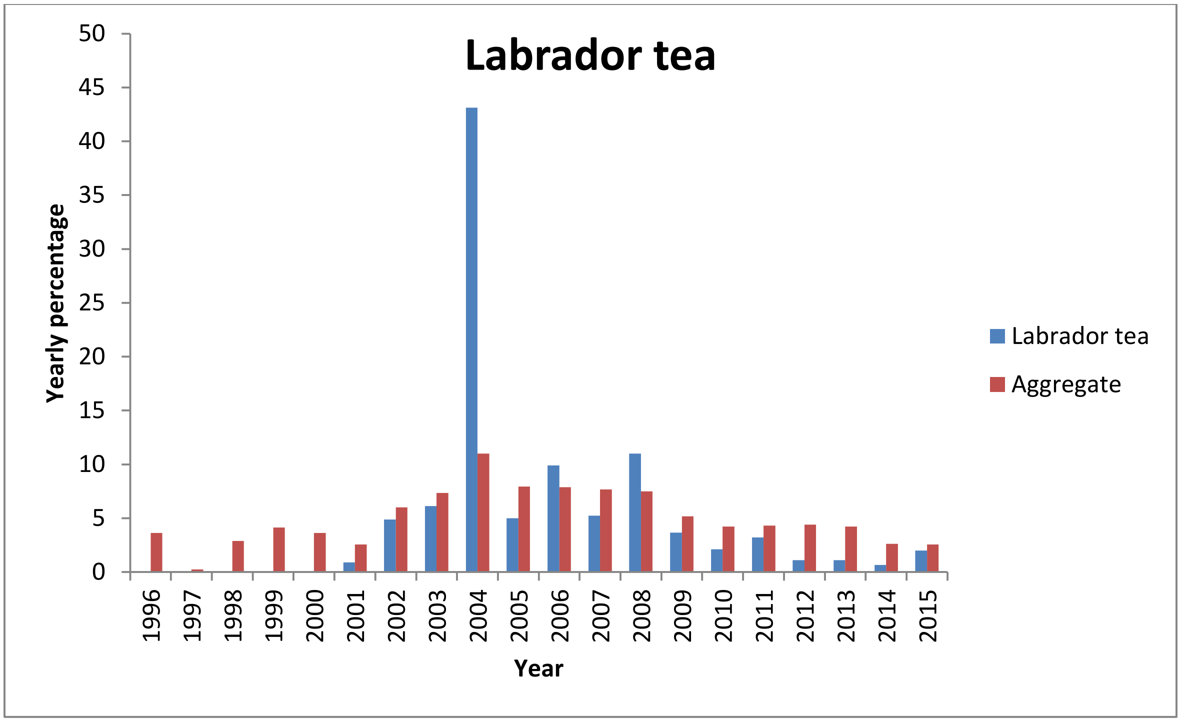

Figure 106: Yearly percentage of the total number of entries for the aggregate and those for Labrador tea.

Labrador tea is a rather abundant species as well with 916 entries prior to 2016. It may not be the best species for analysis though as over 40% of all its entries were in 2004. As such it may be that this species is one of the primary contributors to the spike for the aggregate, which may make it necessary to remove this species from the calculation.

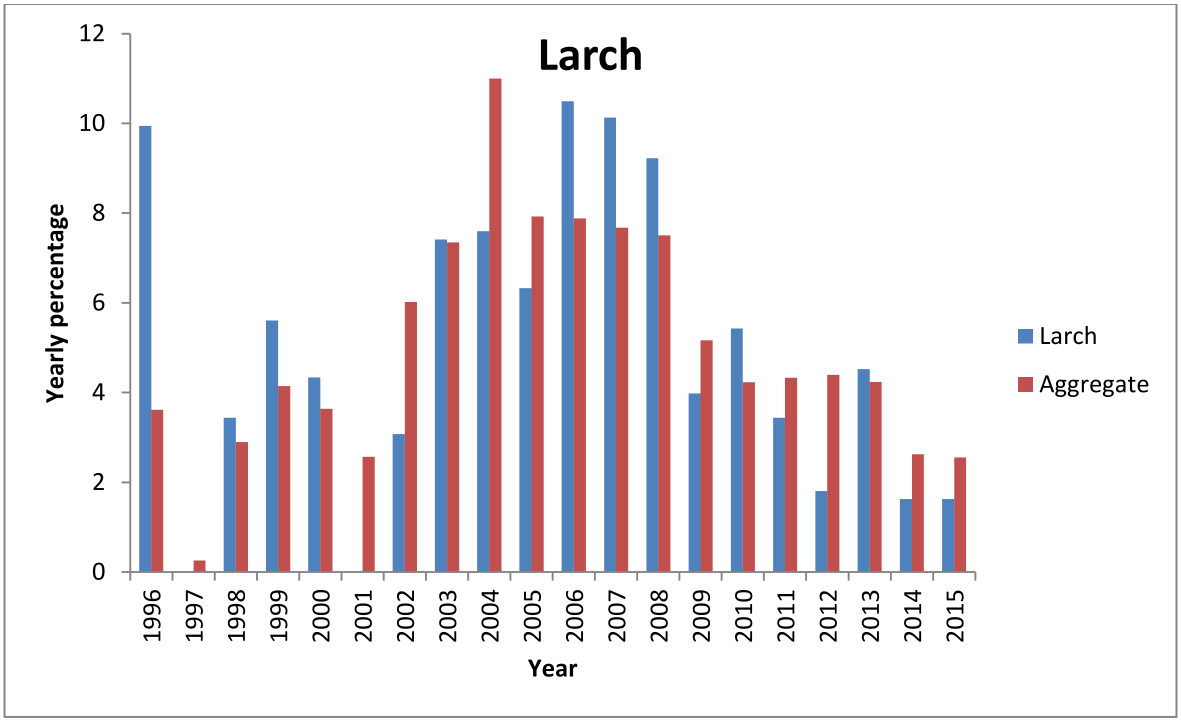

Figure 107: Yearly percentage of the total number of entries for the aggregate and those for Larch.

The number of entries for this species is 552, which is moderately good for analysis. It follows the aggregate very loosely however, especially with regards to the proportion of its entries in 1996.

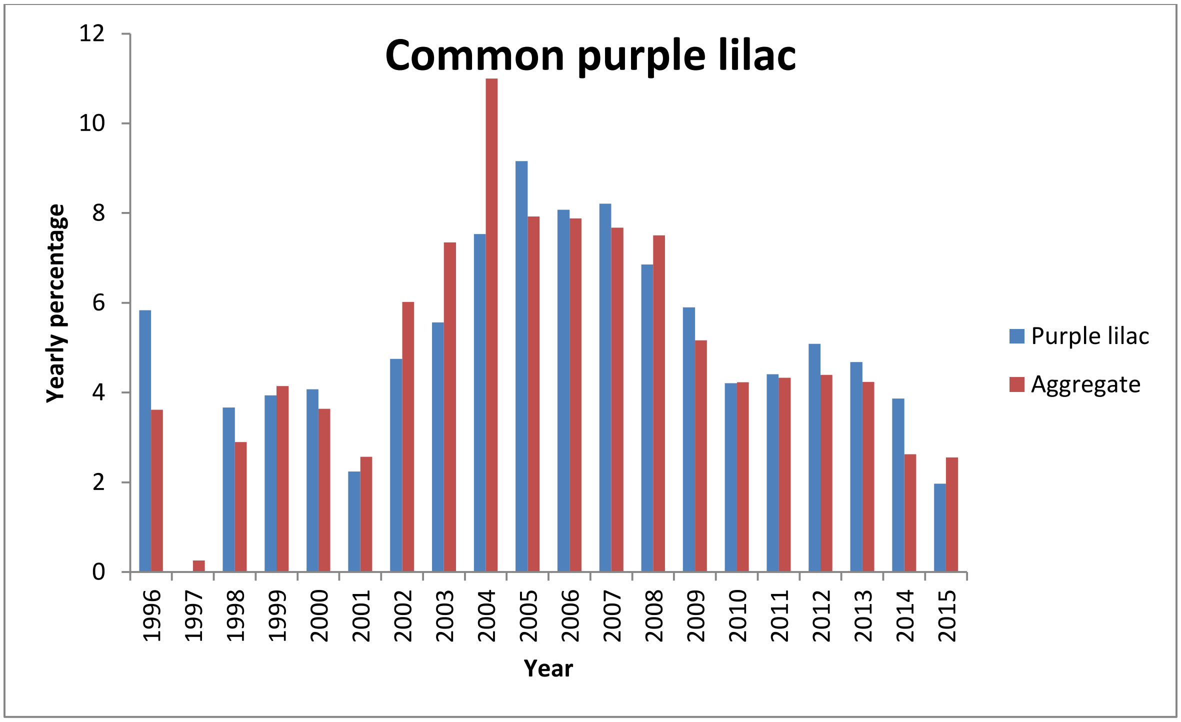

Figure 108: Yearly percentage of the total number of entries for the aggregate and those for Common purple lilac.

This species has an extremely high number of entries at 1482. Being an introduced and often planted species in human-occupied lands it is easy to find, and as such may serve, alongside the Dandelion, to isolate the values resulting from greater participation. The only years where it diverges strongly from the aggregate are 1996 and 2004.

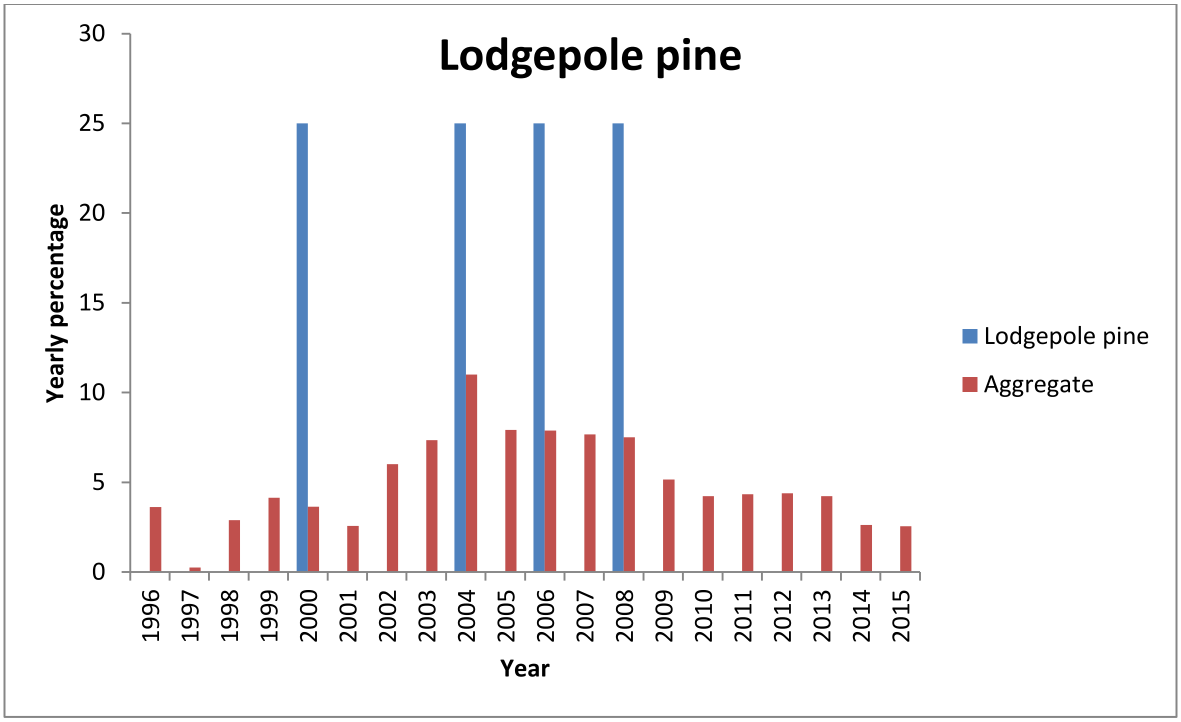

Figure 109: Yearly percentage of the total number of entries for the aggregate and those for Lodgepole pine.

This species only has 4 entries.

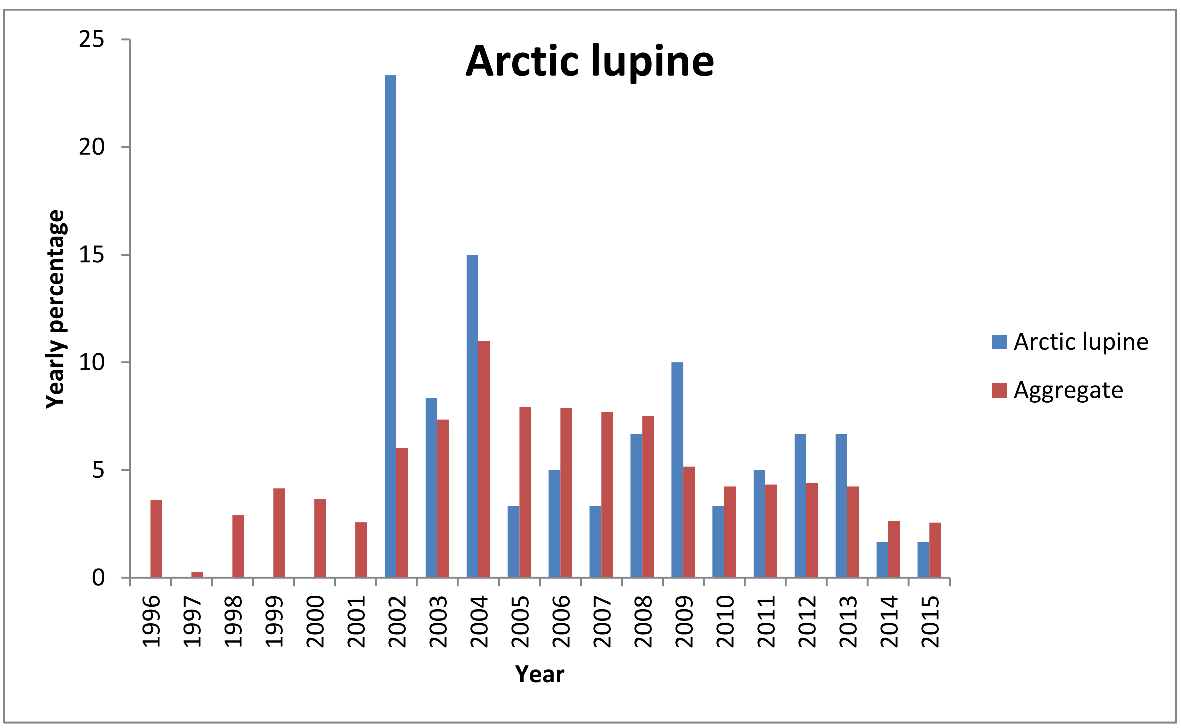

Figure 110: Yearly percentage of the total number of entries for the aggregate and those for Arctic lupine.

This species has only 60 entries.

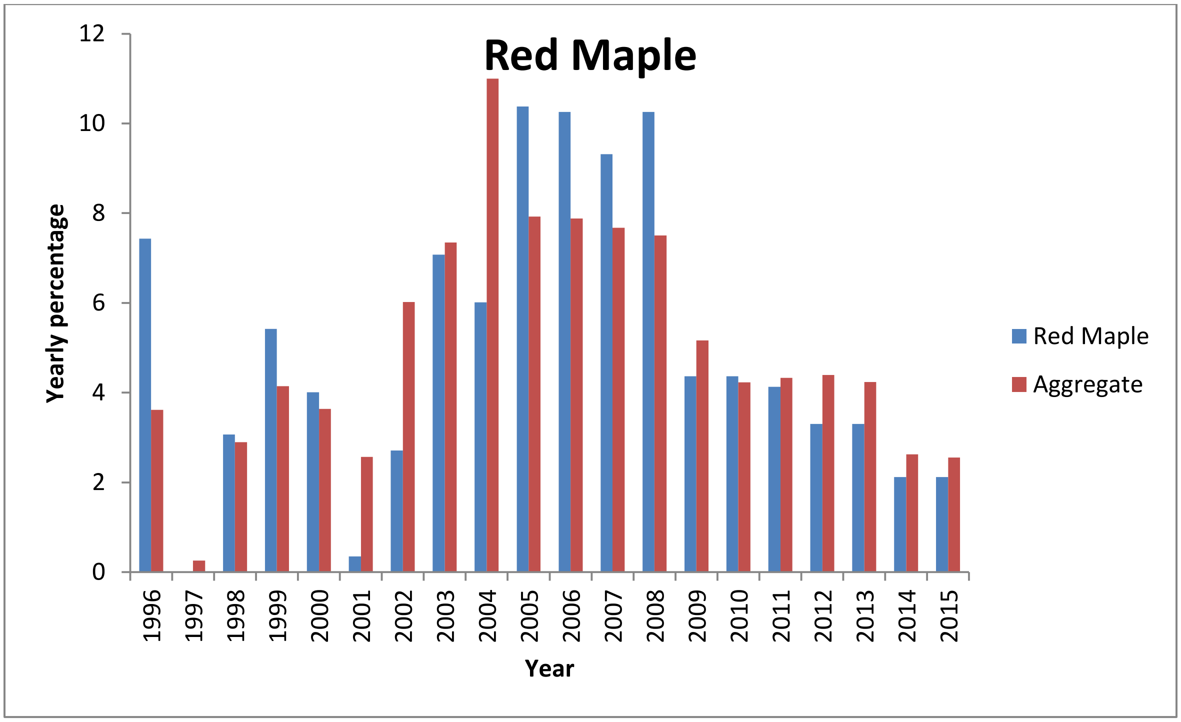

Figure 111: Yearly percentage of the total number of entries for the aggregate and those for Red Maple.

There are 855 entries of Red Maple prior to 2016, which is a very good number. It diverges from the aggregate during several years, notably 2002, 2004, and from 2005 to 2008 by a similar margin.

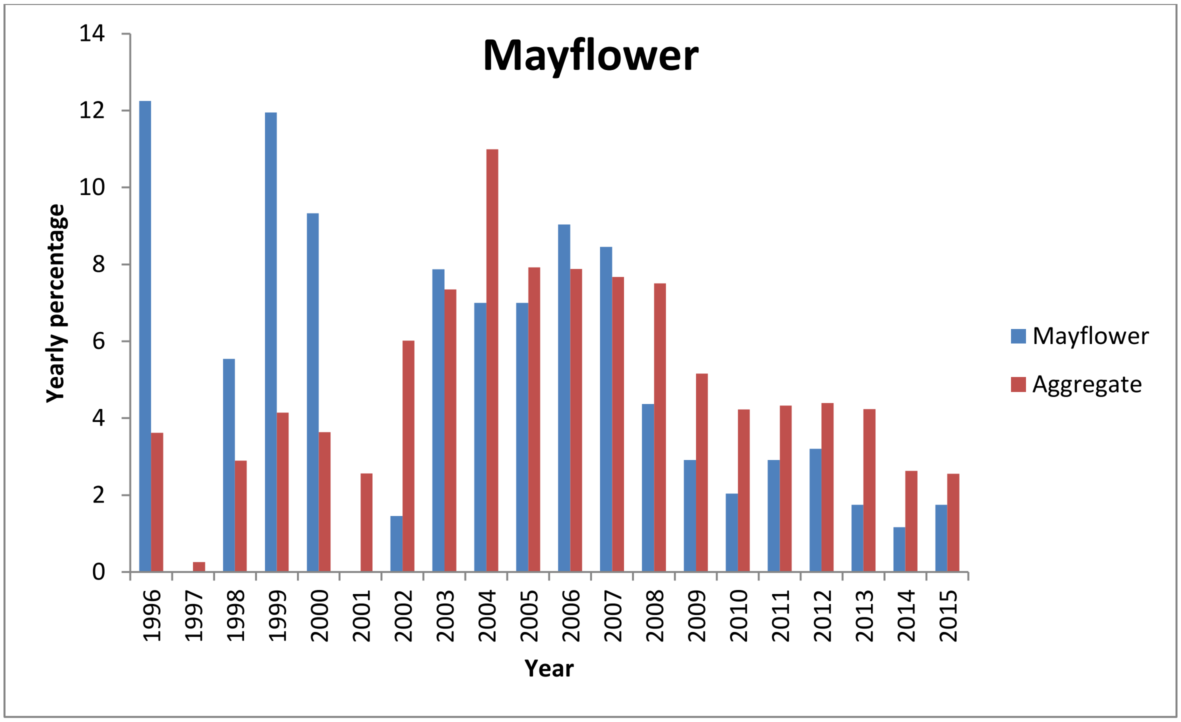

Figure 112: Yearly percentage of the total number of entries for the aggregate and those for Mayflower.

The number of mayflower entries, at 344, is a little too low for analysis, especially considering how many of its entries occurred in 1996, 1999 and 2000.

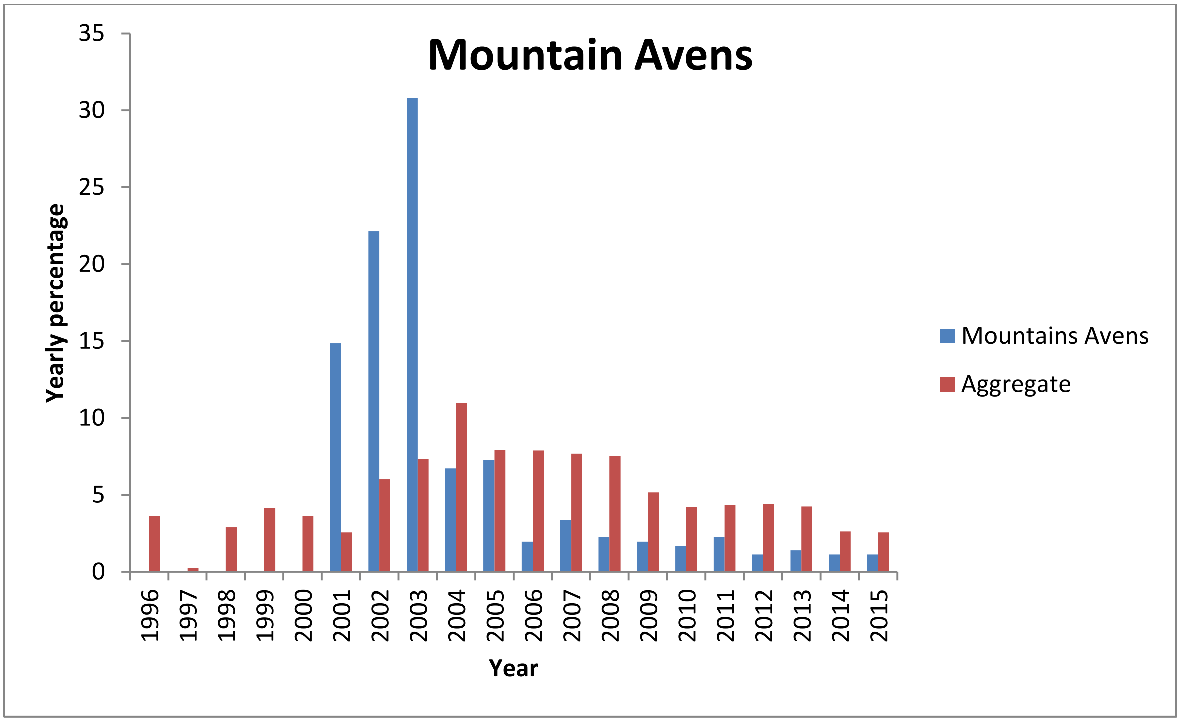

Figure 113: Yearly percentage of the total number of entries for the aggregate and those for Mountain avens.

This species has 359 entries, but it is especially inconvenient for analysis given that the crushing majority of its entries were between 2001 and 2003.

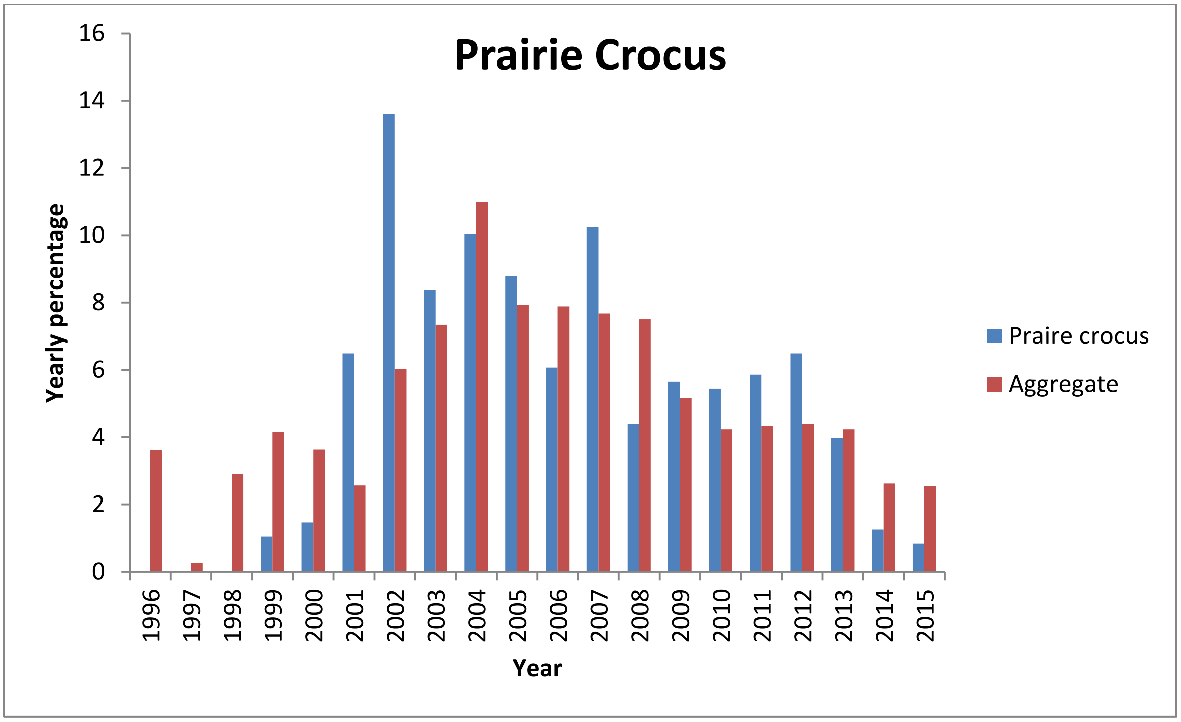

Figure 114: Yearly percentage of the total number of entries for the aggregate and those for Prairie crocus.

There is a good number of entries for this species (493). It has unusual peaks in 2001 and 2002, and in general diverges noticeably (but not as distinctly) from the aggregate for most years.

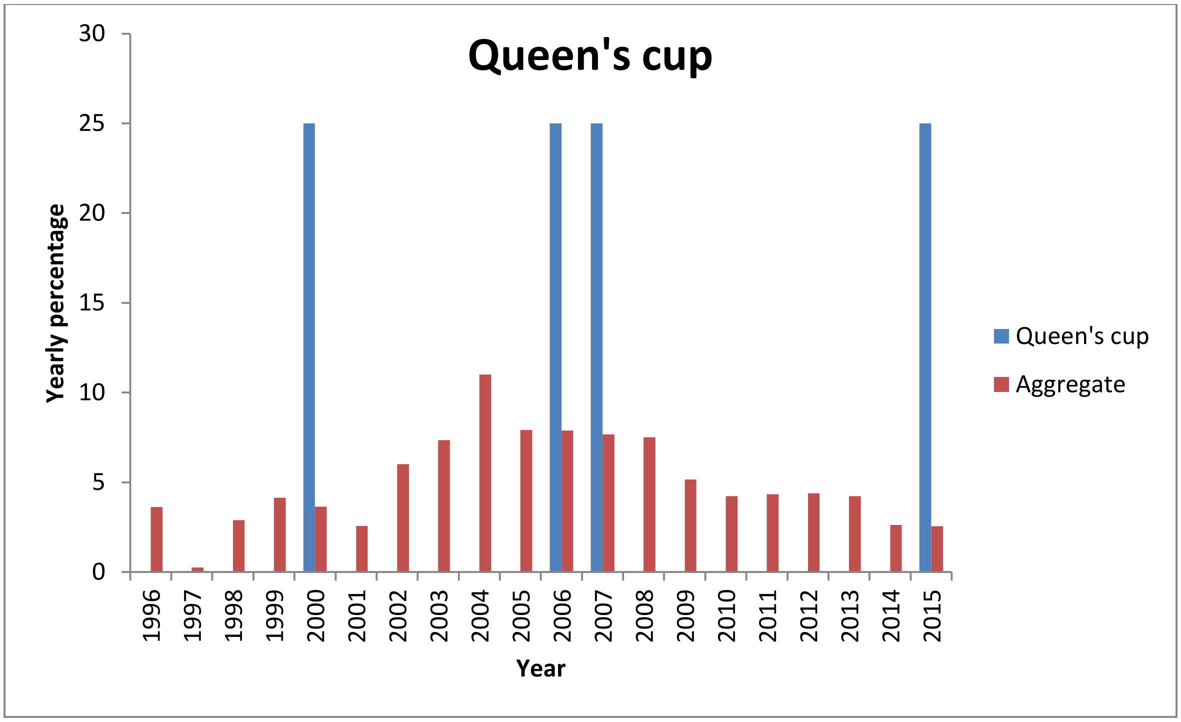

Figure 115: Yearly percentage of the total number of entries for the aggregate and those for Queen's cup.

There are only 4 entries for this species.

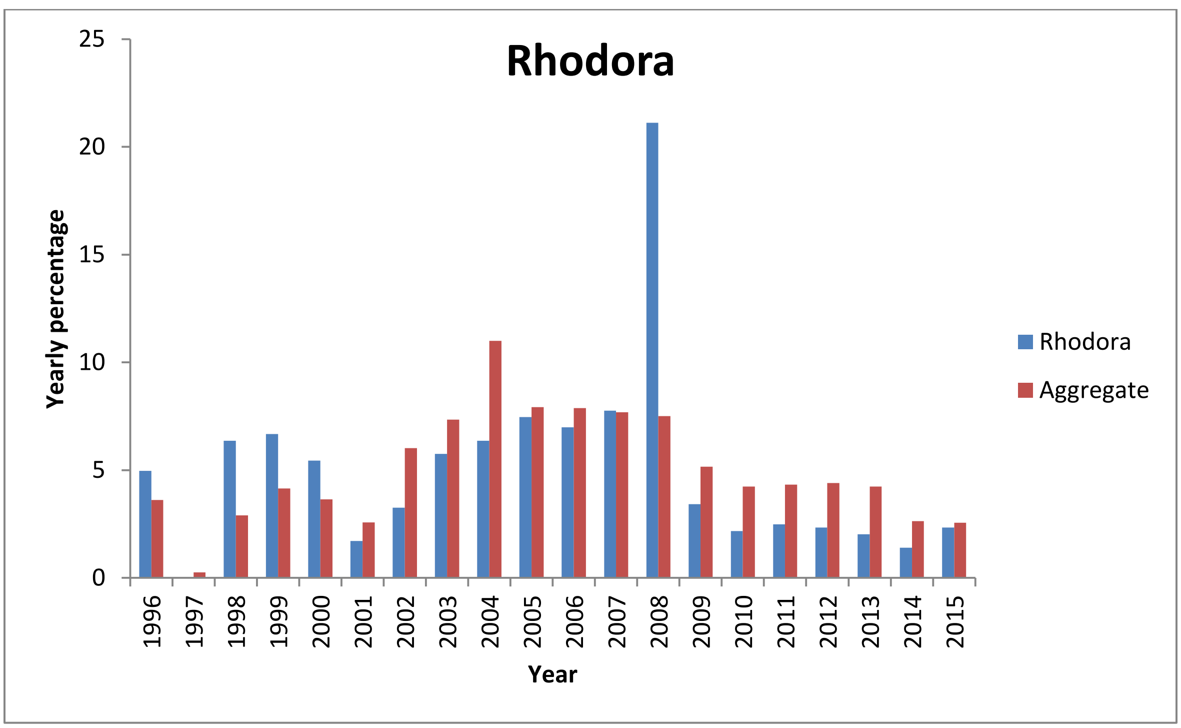

Figure 116: Yearly percentage of the total number of entries for the aggregate and those for Rhodora.

This species has a very good number of entries at 648. Interestingly it follows the aggregate very closely for the most part except for a spike in 2008.

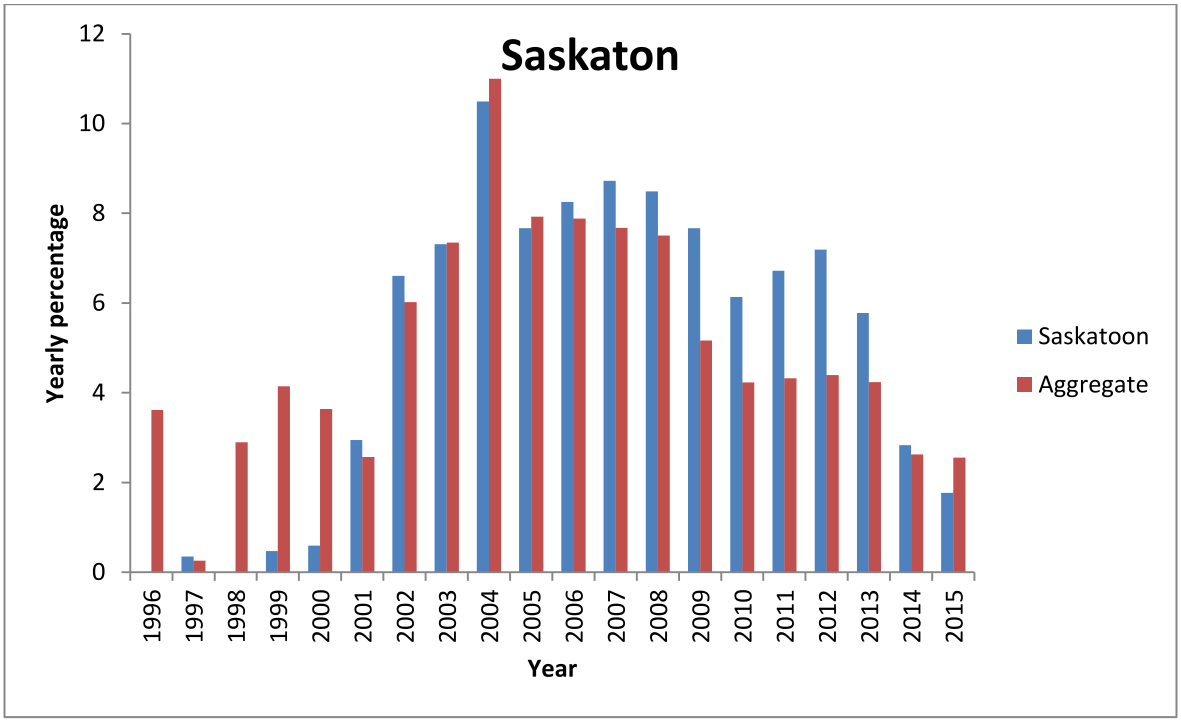

Figure 117: Yearly percentage of the total number of entries for the aggregate and those for Saskatoon.

The Saskatoon is another good species for analysis with 877 entries. This species is notable for having percentages noticeably above the aggregate from 2009 to 2013, otherwise it follows the aggregate from 2001 onwards (with few entries at all before this year).

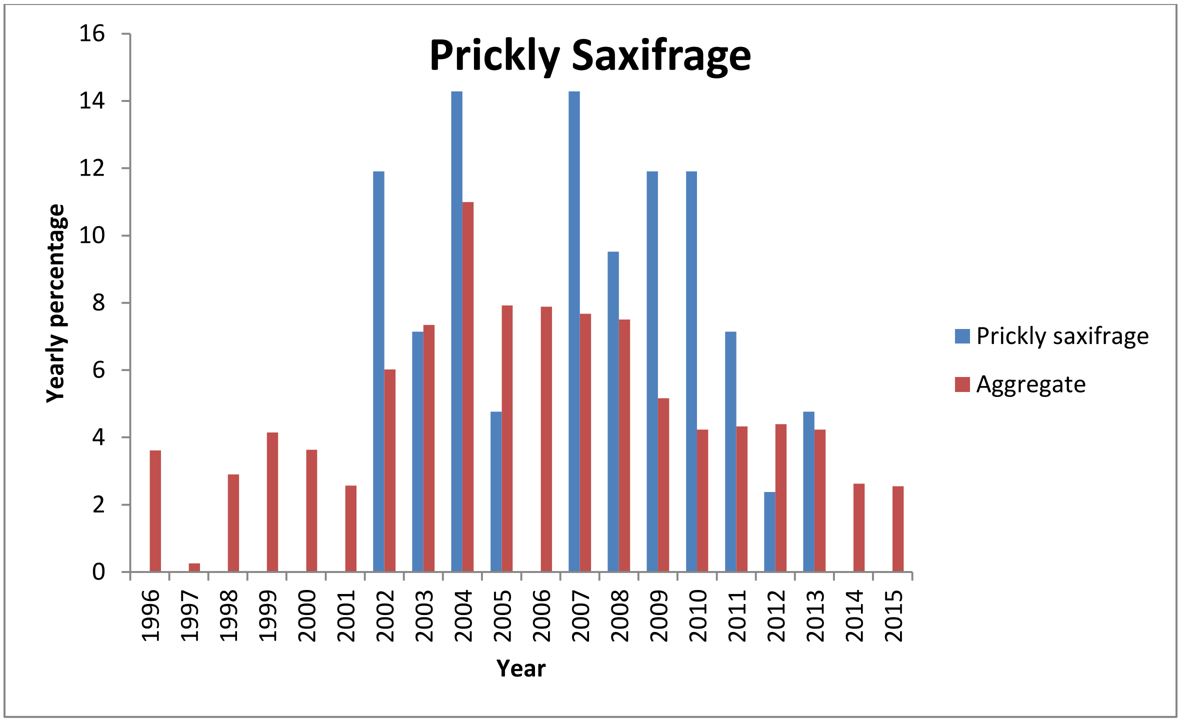

Figure 118: Yearly percentage of the total number of entries for the aggregate and those for Prickly saxifrage.

This species only has 42 entries.

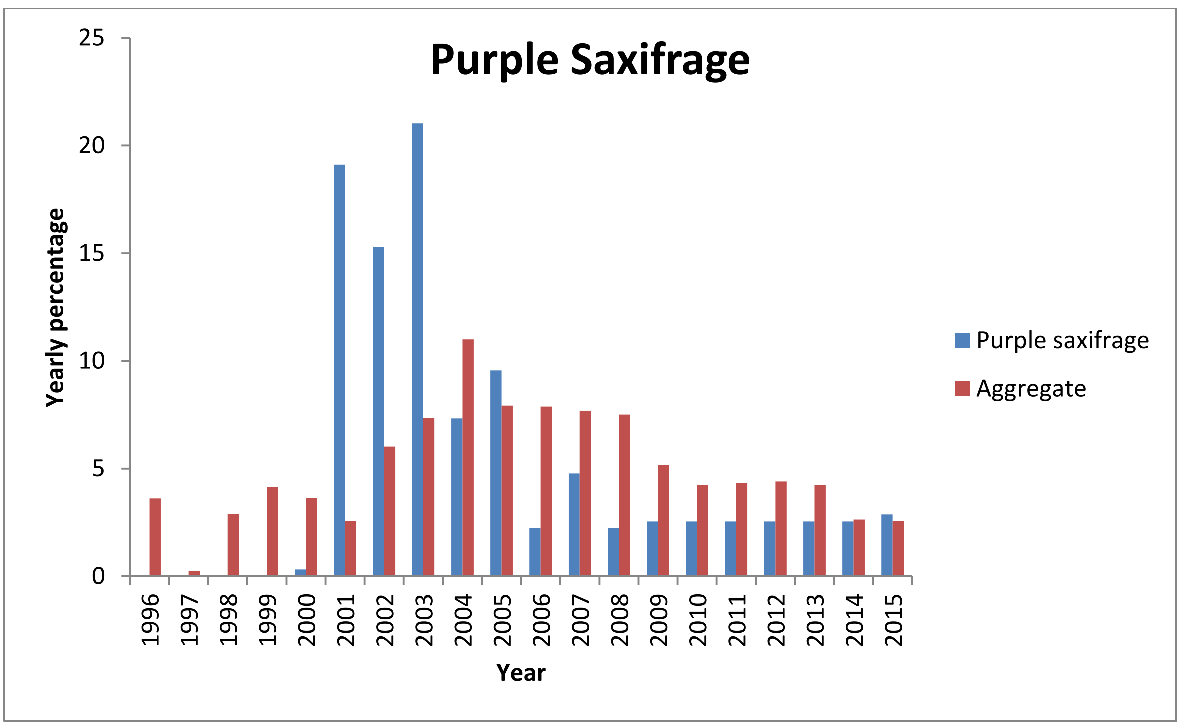

Figure 119: Yearly percentage of the total number of entries for the aggregate and those for Purple saxifrage.

With 315 entries, this species is far from ideal for analysis. The percentage of entries occurring between 2001 and 2003 casts doubt on the usefulness of this species' data.

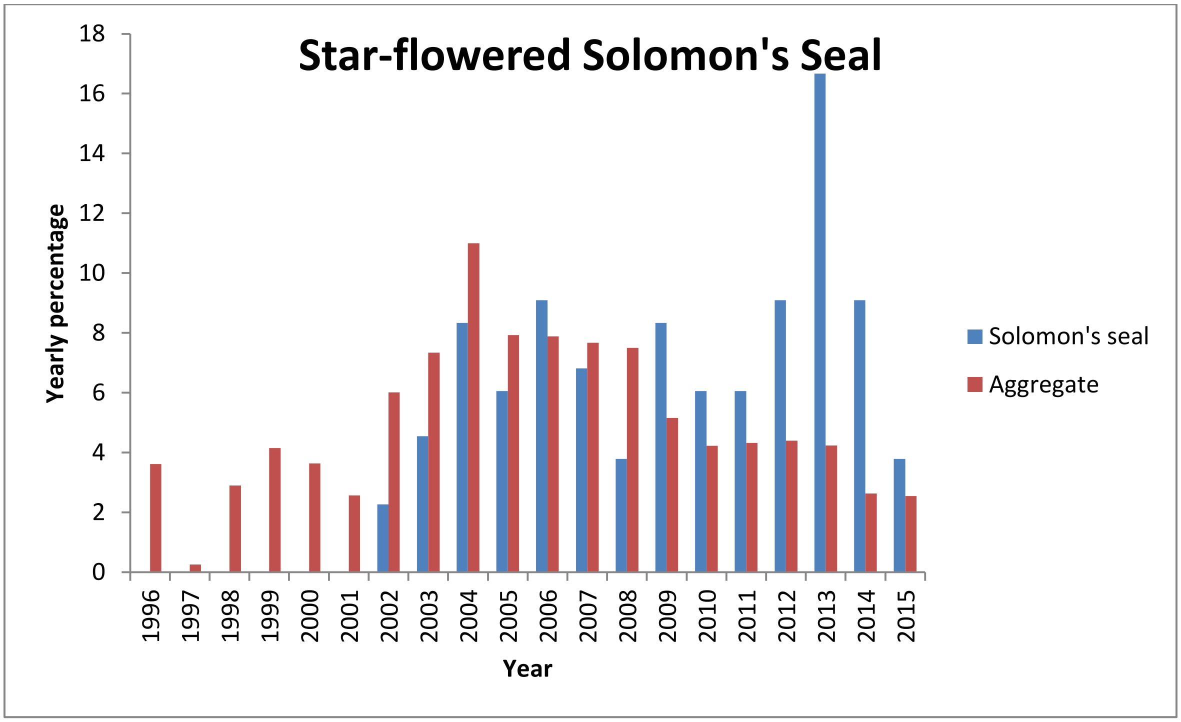

Figure 120: Yearly percentage of the total number of entries for the aggregate and those for Star-flowered Solomon's seal.

At 132, the entry count for this species is too low to be of much use for analysis.

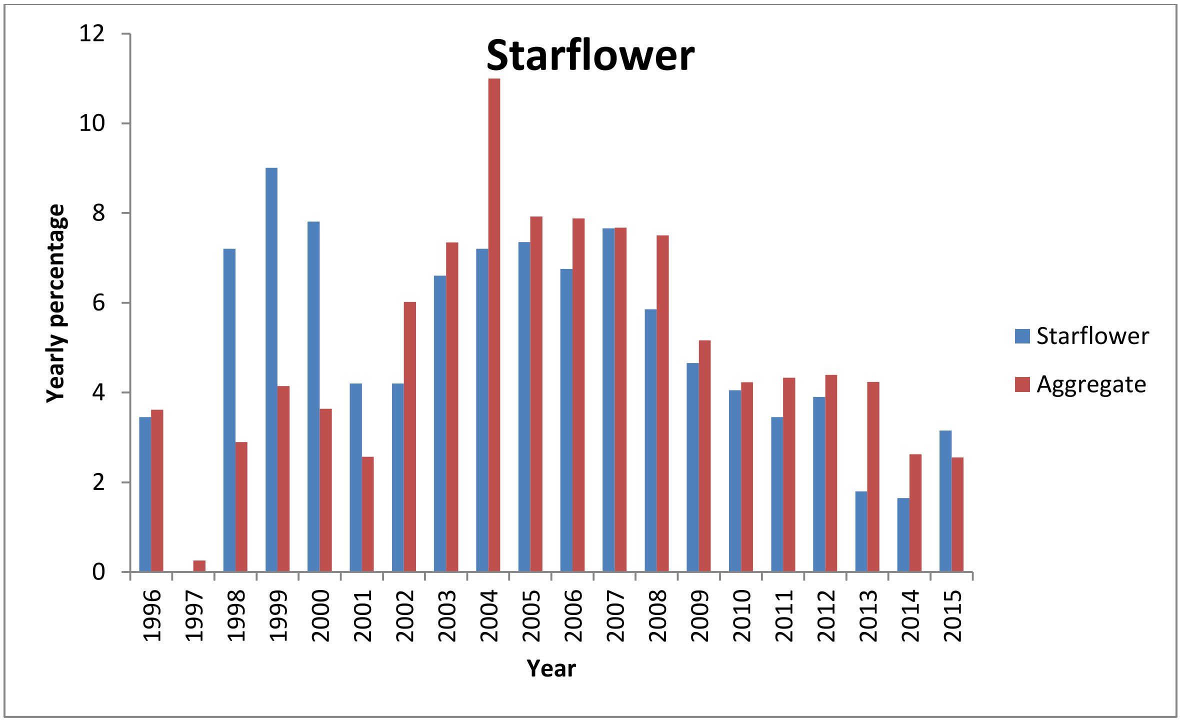

Figure 121: Yearly percentage of the total number of entries for the aggregate and those for Starflower.

While an unfortunate number by some metrics, this species is of good value for analysis with 666 entries before 2016. The years from 1998 to 2000 would be interesting to analyse given how they strongly differ from the aggregate, but retain a similar profile.

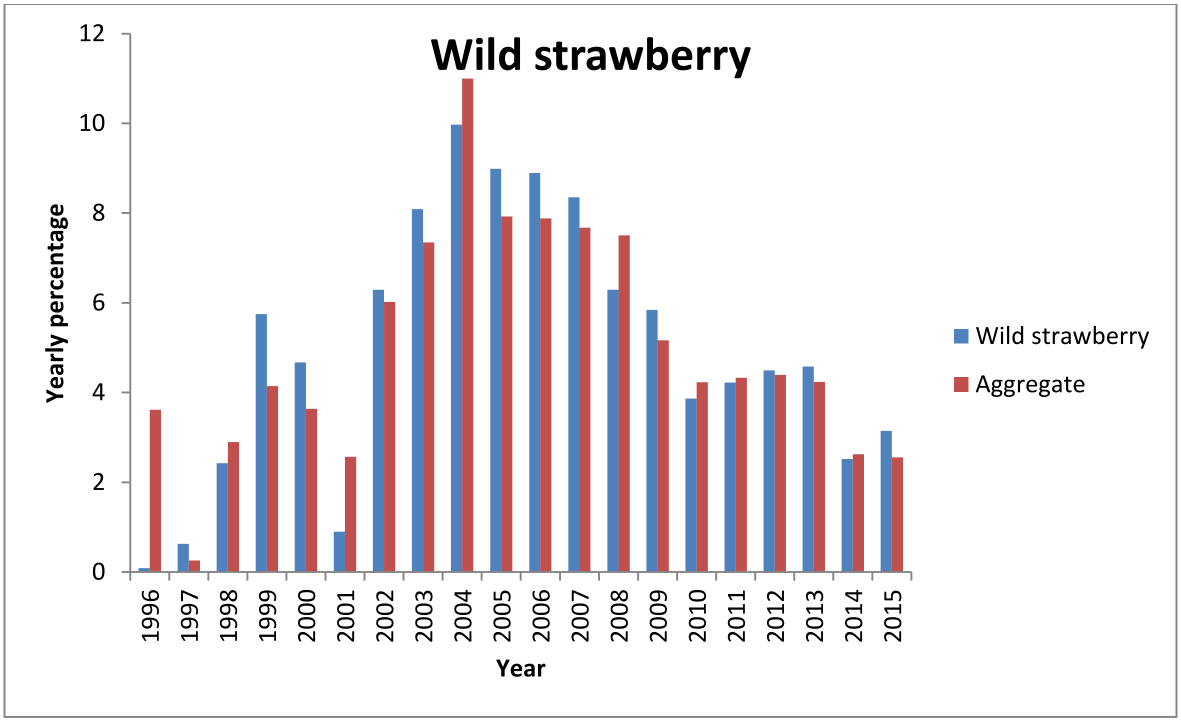

Figure 122: Yearly percentage of the total number of entries for the aggregate and those for Wild Strawberry.

It comes as little surprise that a widely sought-after species would have such a high number of entries (1124). It is also notable in how closely it follows the aggregate in terms of trends, besides the very early years before the official formation of NatureWatch. This species has potential value as a control or baseline to compare against as it is very well known, and as its climate preferences are equally familiar it may also serve as an analog to weather when compared to the aggregate. The possible issue with this species is confusion with strawberries grown as crops or plants originating from "domestic" strawberry seeds dispersed outside of fields.

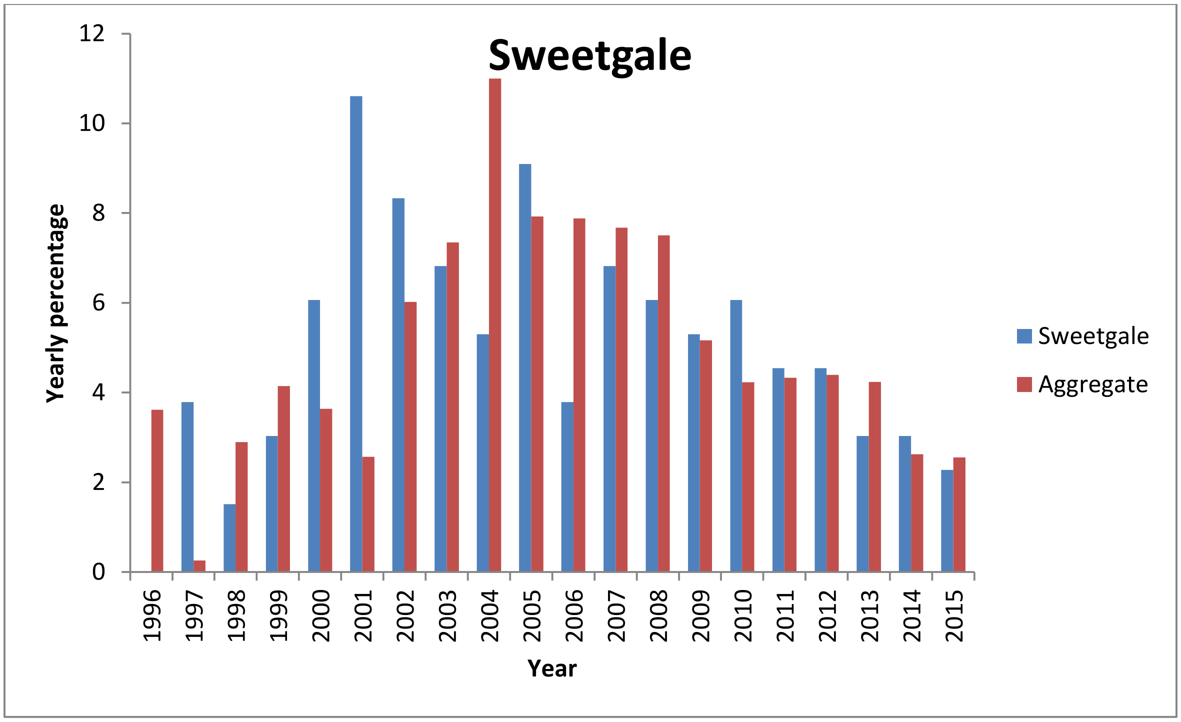

Figure 123: Yearly percentage of the total number of entries for the aggregate and those for Sweetgale.

With 131 entries before 2016, this species would be too small a sample for analysis.

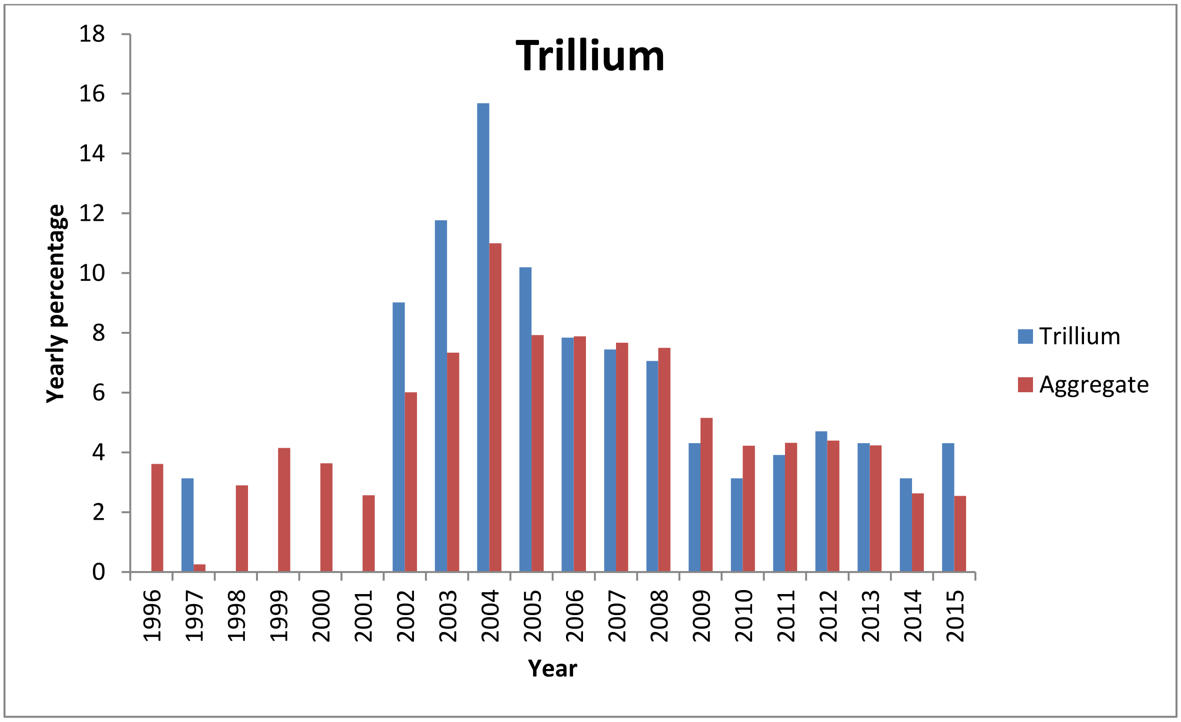

Figure 124: Yearly percentage of the total number of entries for the aggregate and those for Trillium.

Trillium has only 265 entries, which is not enough to be of much use for study. Interestingly though after 2004 its percentages roughly follow the aggregate.

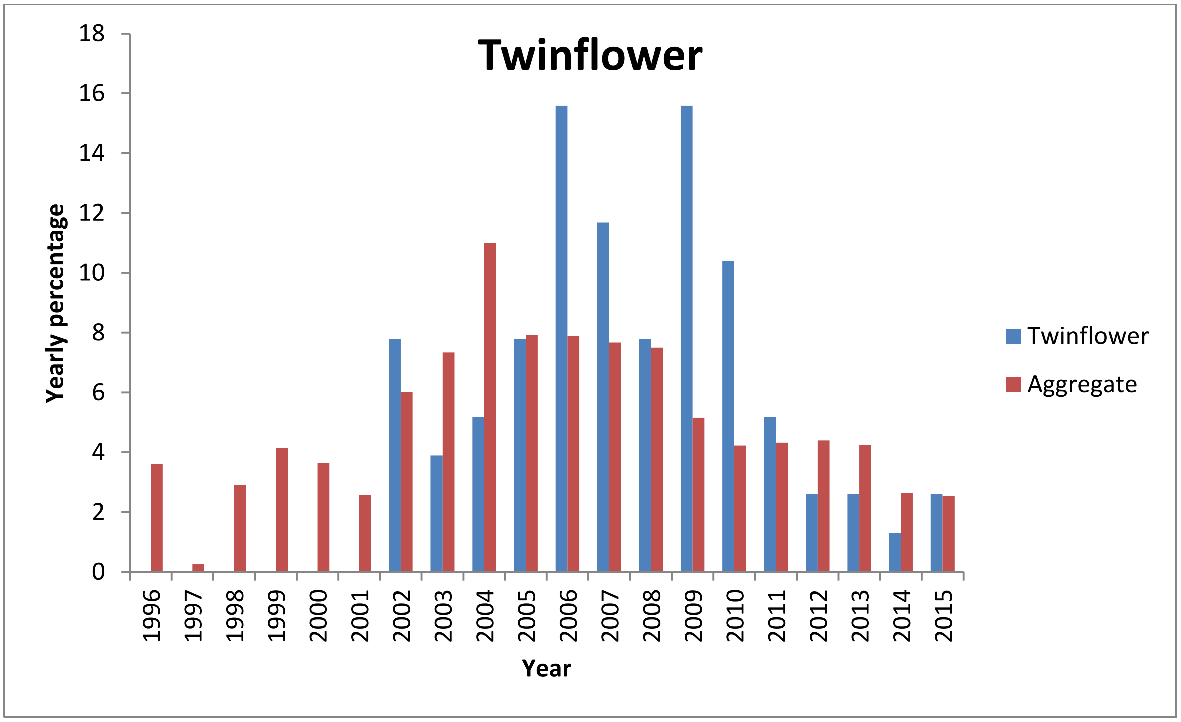

Figure 125: Yearly percentage of the total number of entries for the aggregate and those for Twinflower.

This species only has 77 entries.

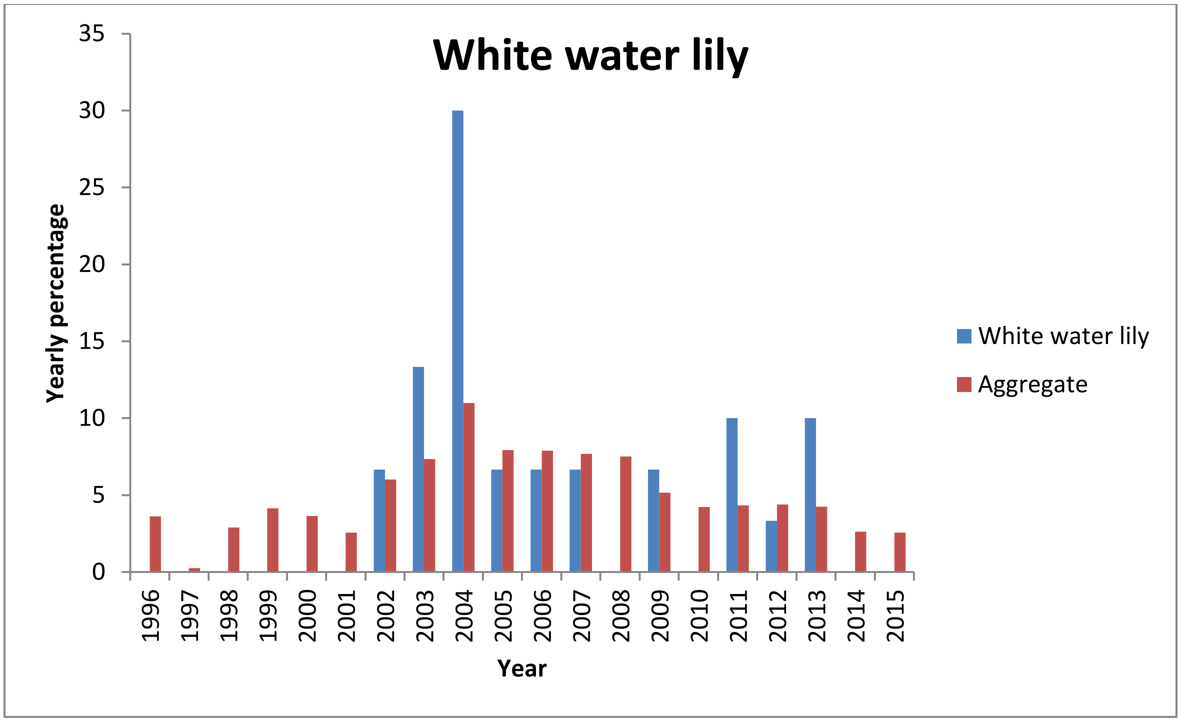

Figure 126: Yearly percentage of the total number of entries for the aggregate and those for White water lily.

This species only has 30 entries.

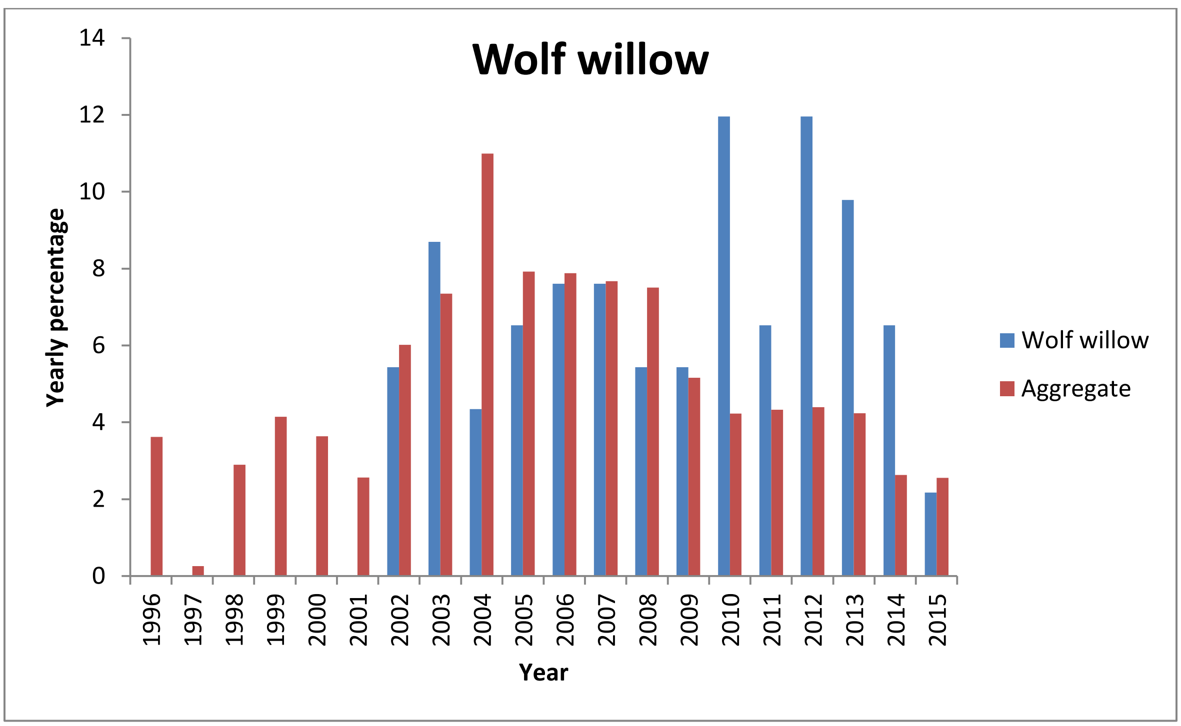

Figure 127: Yearly percentage of the total number of entries for the aggregate and those for Wolf willow.

This species only has 90 entries.

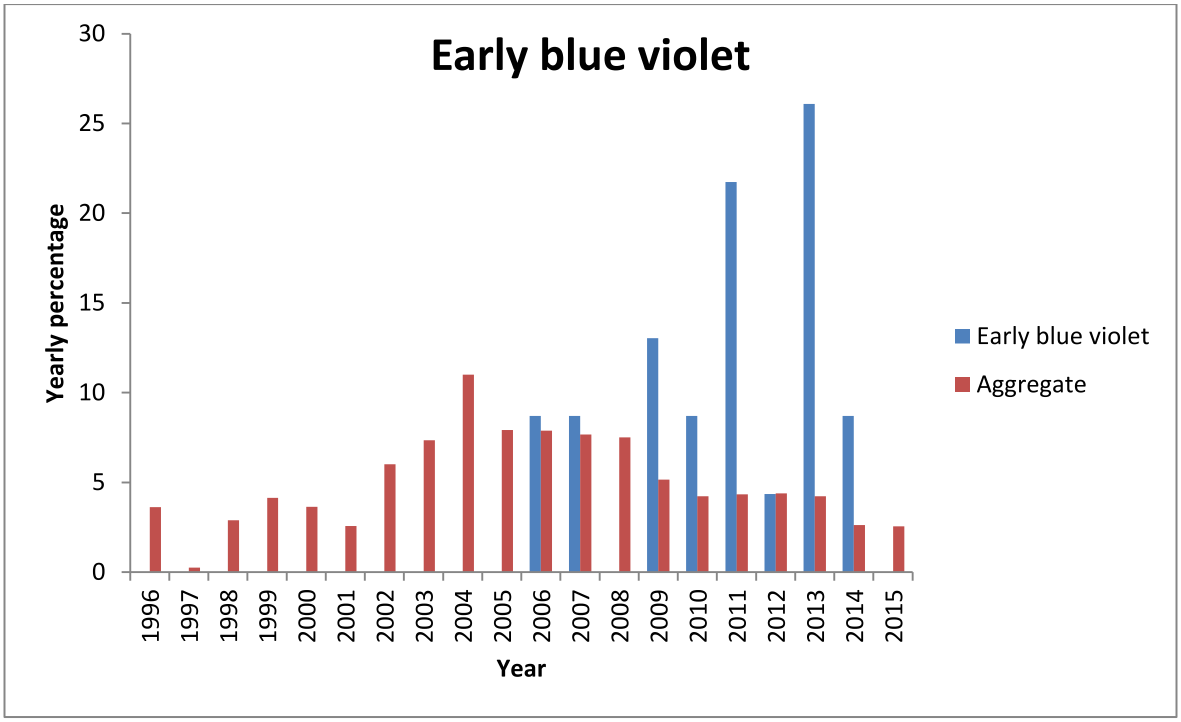

Figure 128: Yearly percentage of the total number of entries for the aggregate and those for Early blue violet.

This species only has 23 entries.

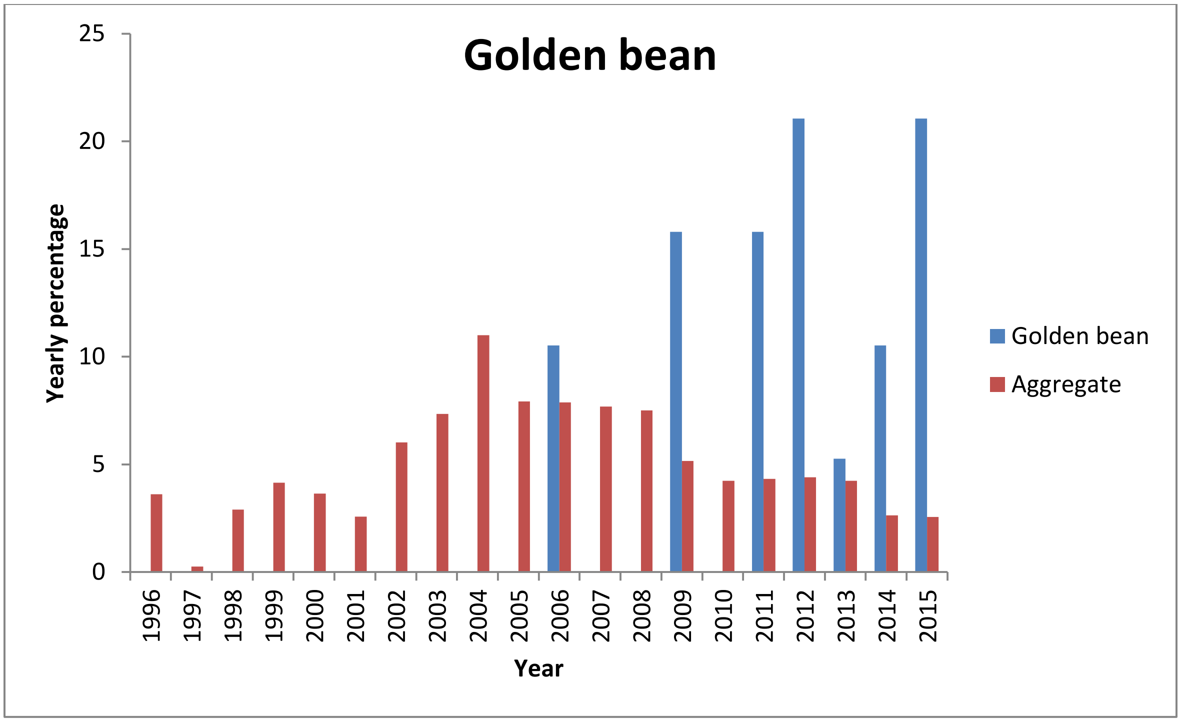

Figure 129: Yearly percentage of the total number of entries for the aggregate and those for Golden bean.

This species only has 16 entries.

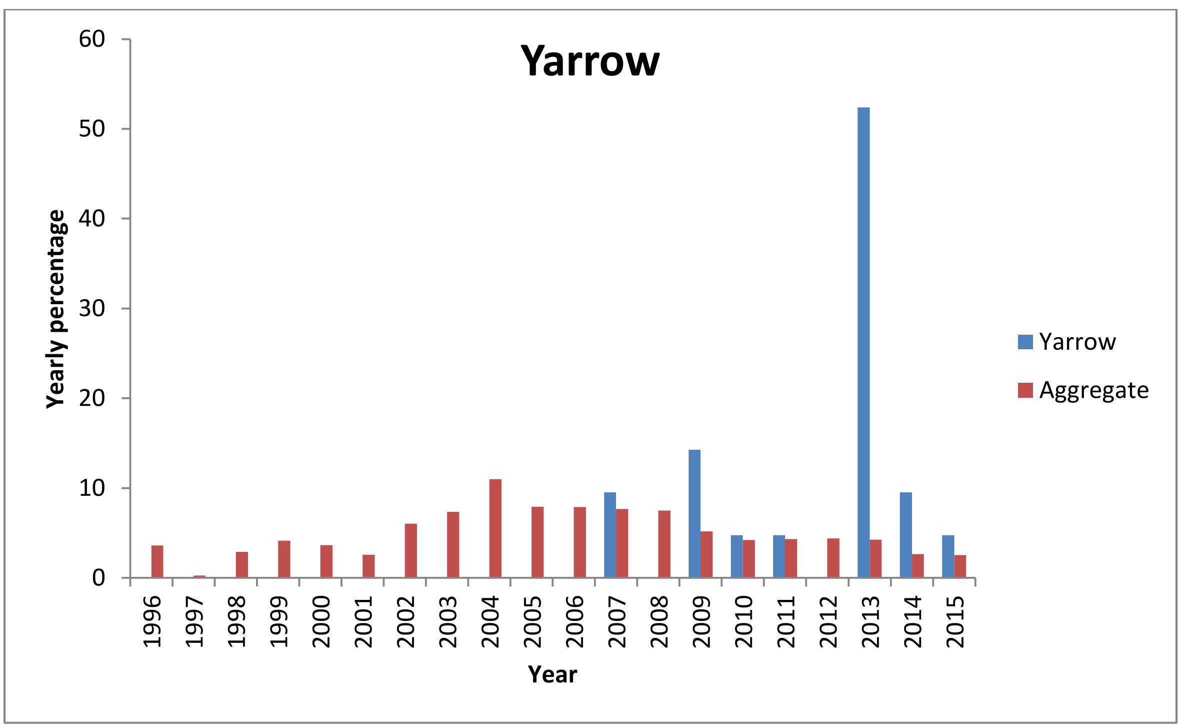

Figure 130: Yearly percentage of the total number of entries for the aggregate and those for Yarrow.

This species only has 23 entries.

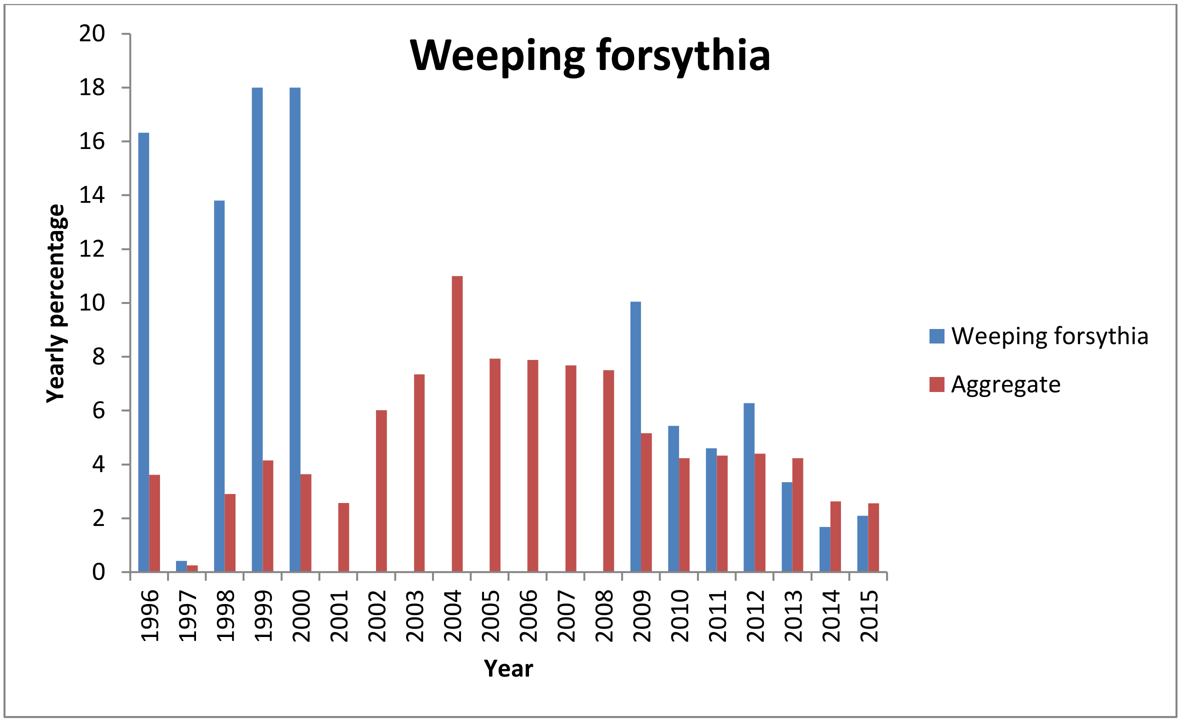

Figure 131: Yearly percentage of the total number of entries for the aggregate and those for Weeping forsythia.

With 241 entries, this species is not particularly good for analysis, especially when we consider the large gap in years with no entries at all.

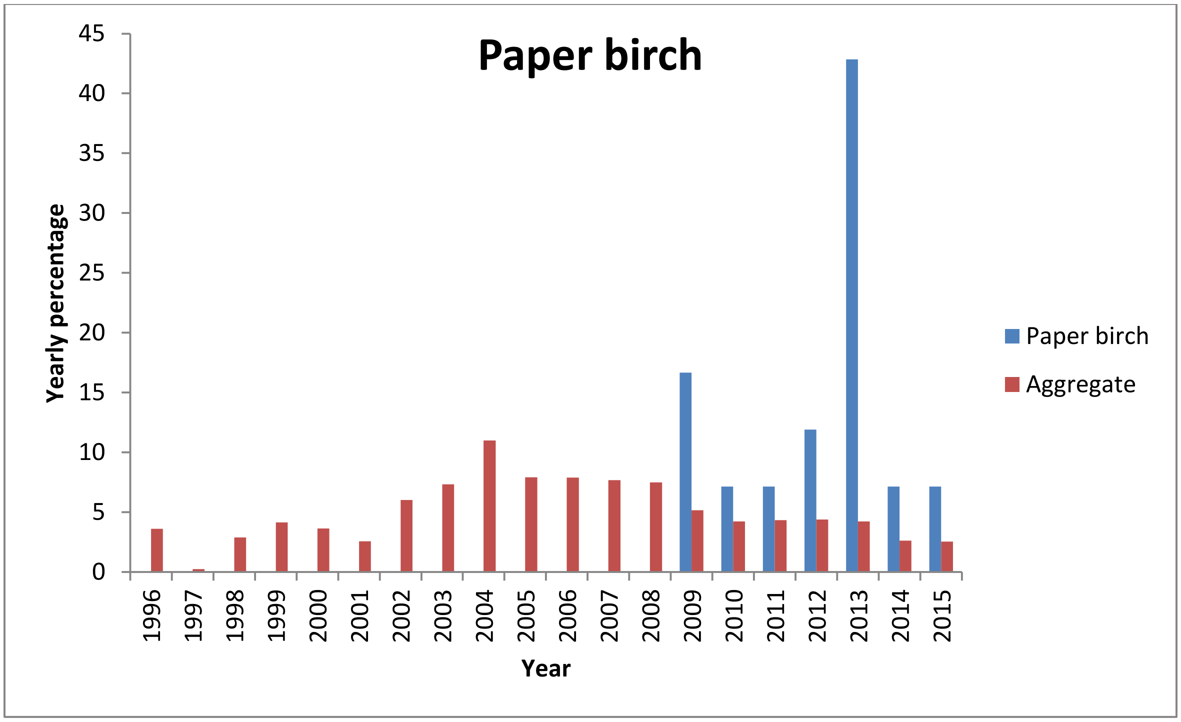

Figure 132: Yearly percentage of the total number of entries for the aggregate and those for Paper birch.

Surprisingly for this author, Paper Birch only has 43 entries.

Appendix B-3: Plantwatch spatial analysis

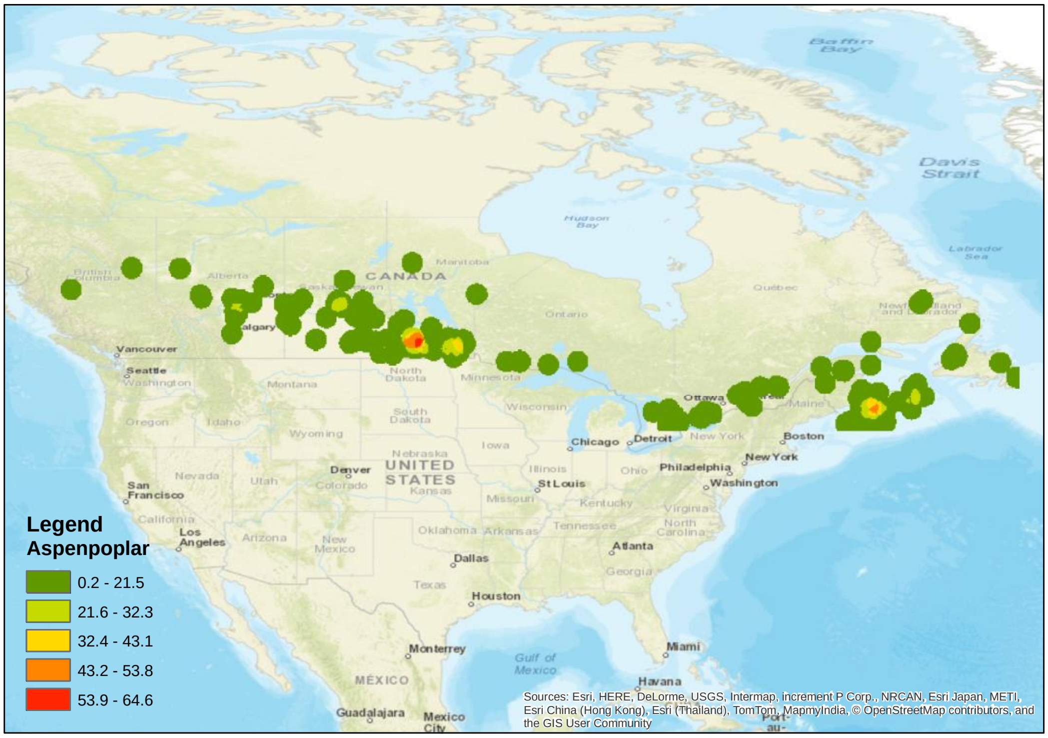

Figure 133: Map of the entry density for Aspen poplar.

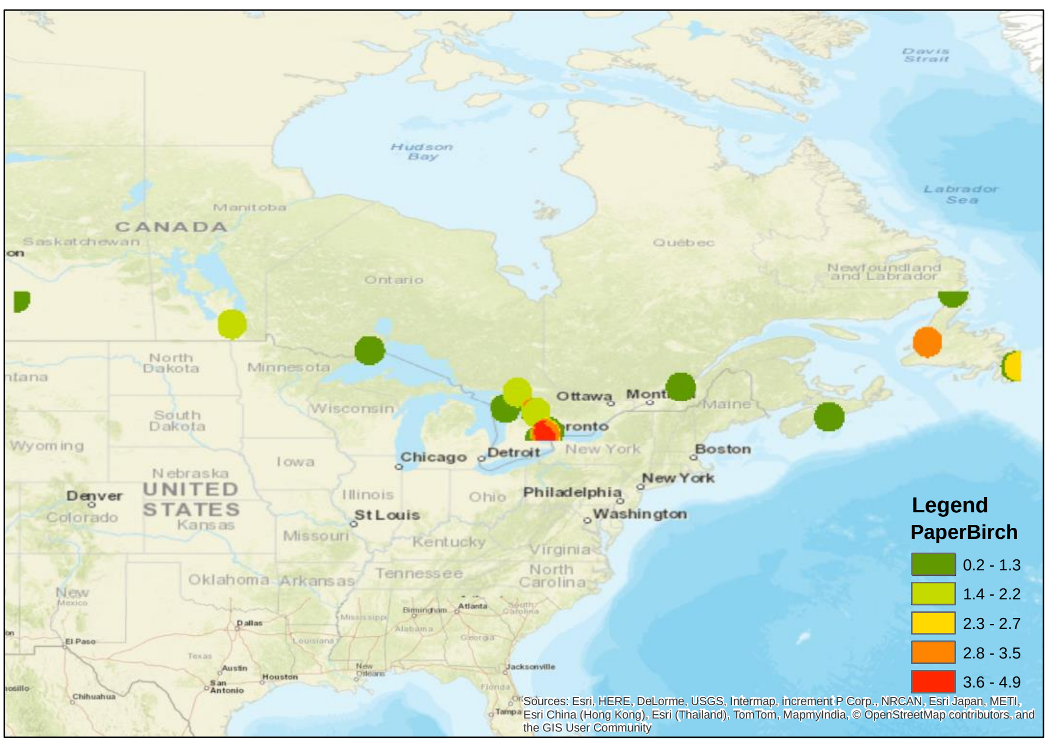

Figure 134: Entry density for Paper birch for Plantwatch.

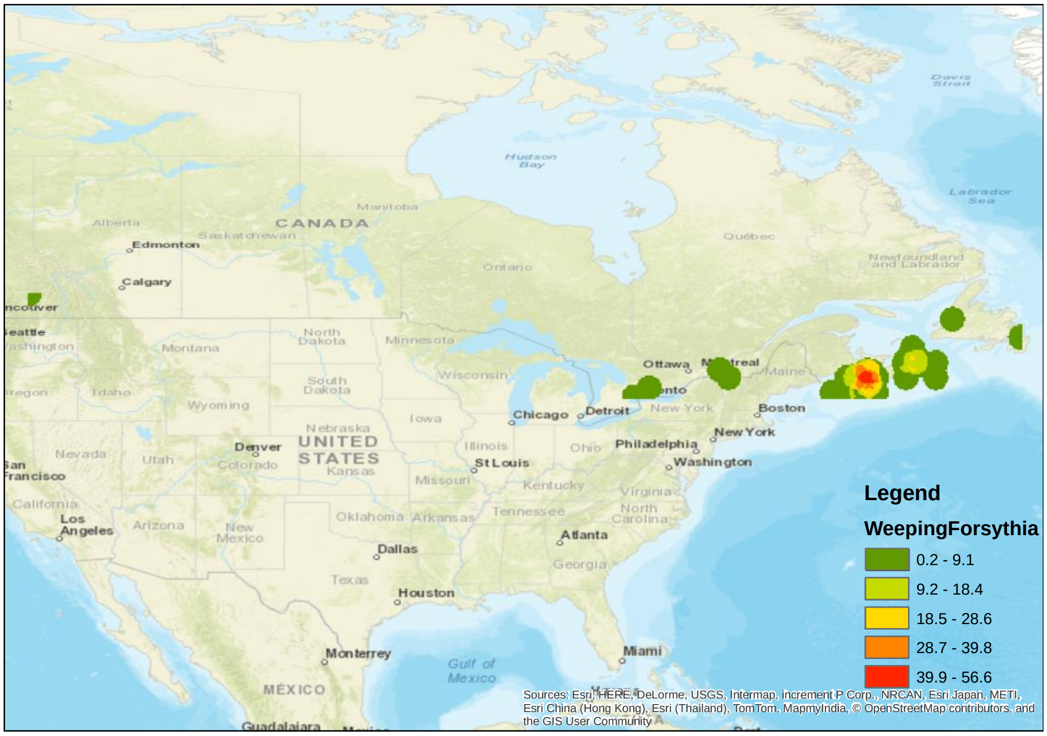

Figure 135: Weeping forsythia entry density map.

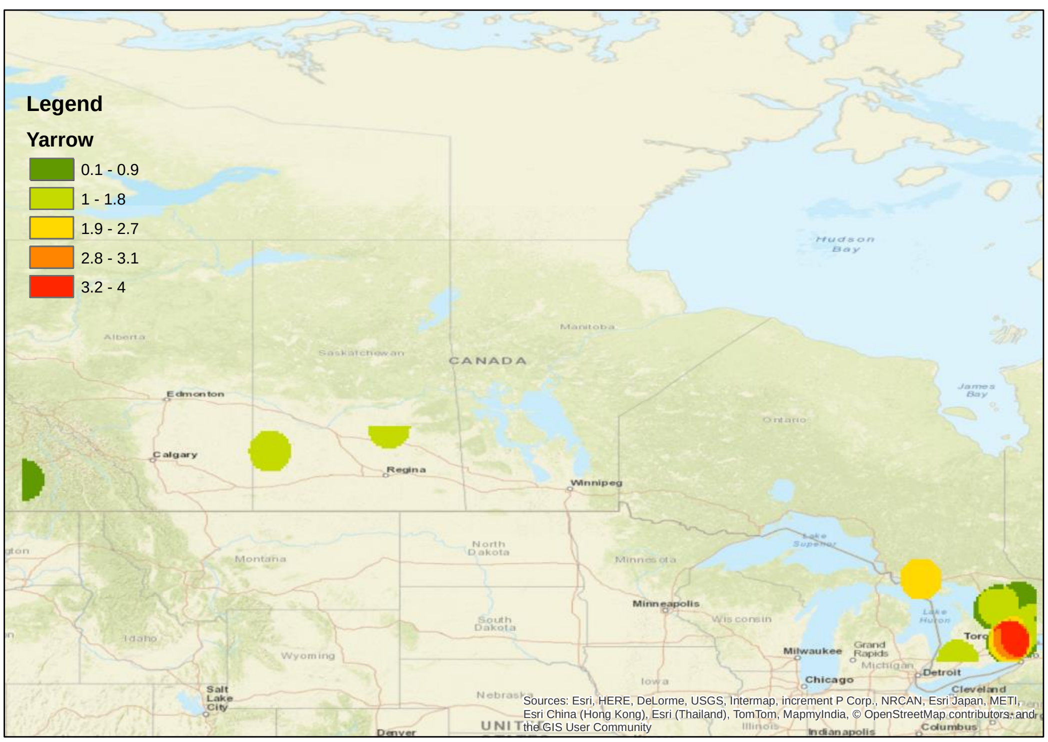

Figure 136: Entry density map for Yarrow.

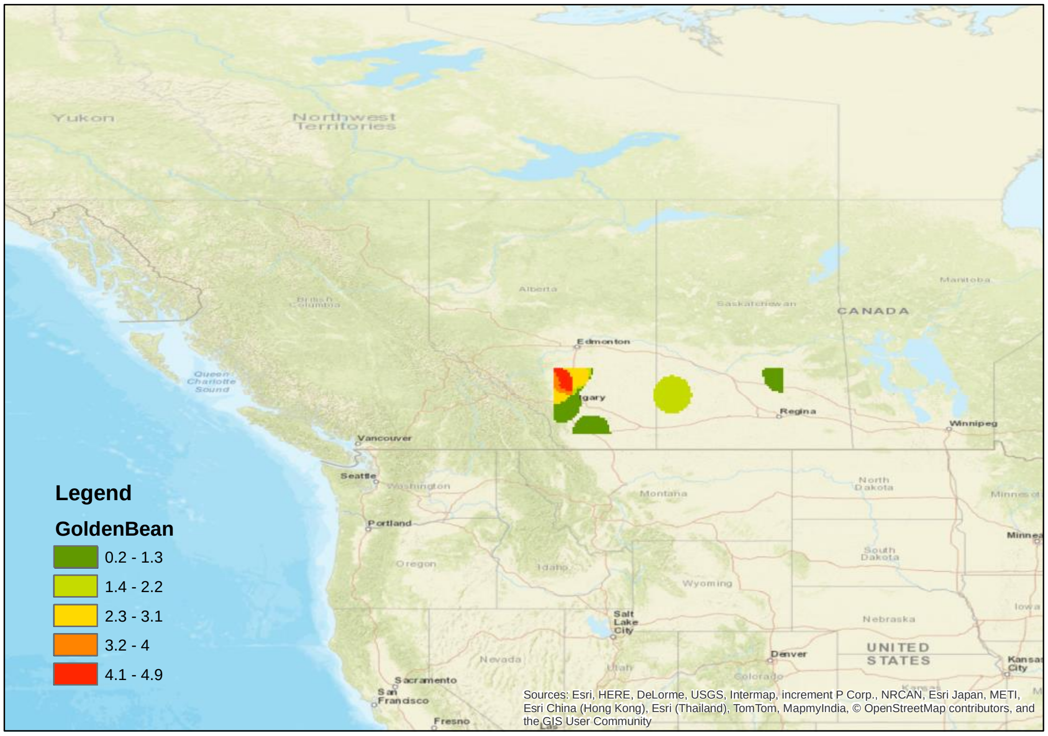

Figure 137: Golden bean entry density map.

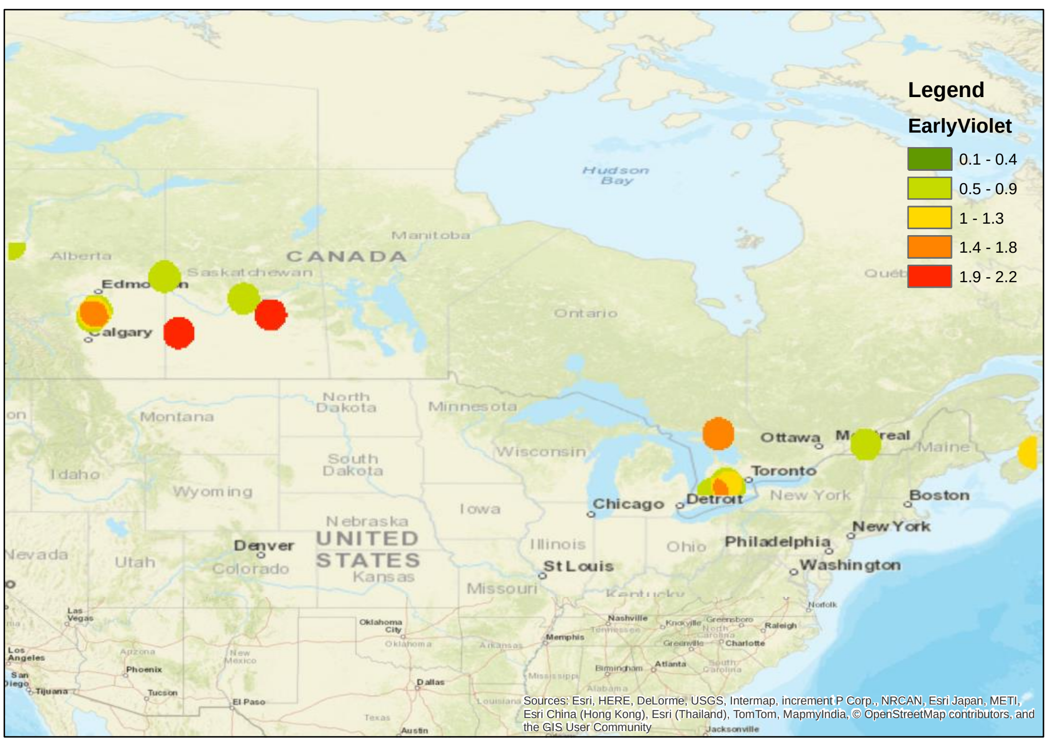

Figure 138: Early violet entry density map.

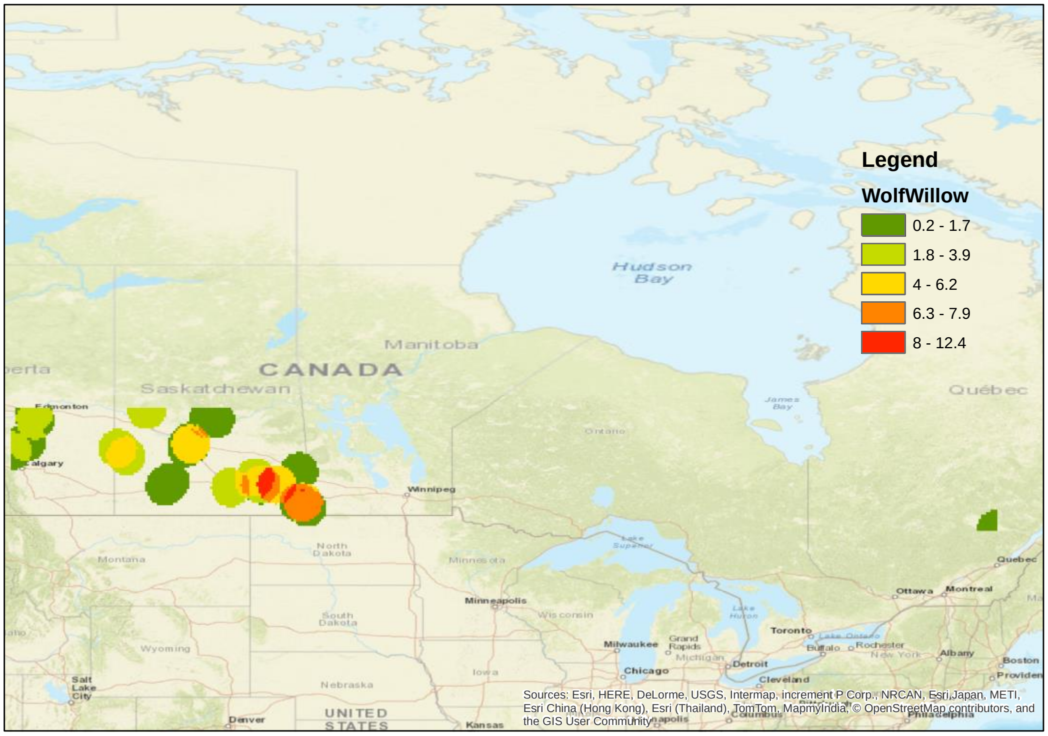

Figure 139: Entry density map for Wolf willow.

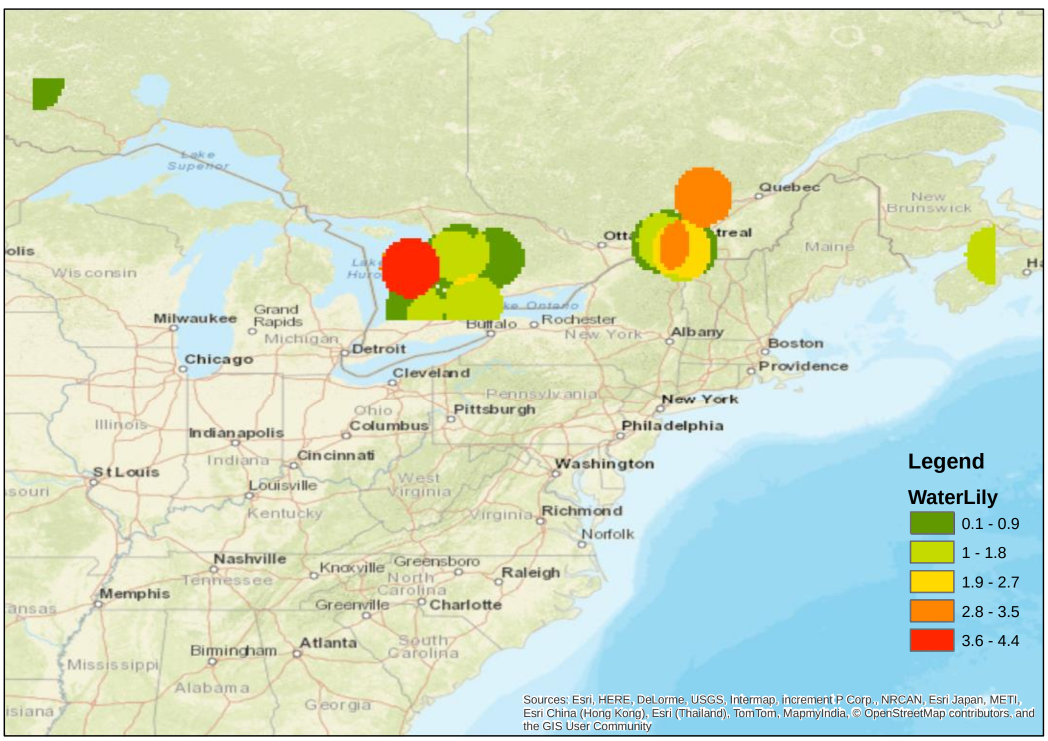

Figure 140: Water lily entry density map.

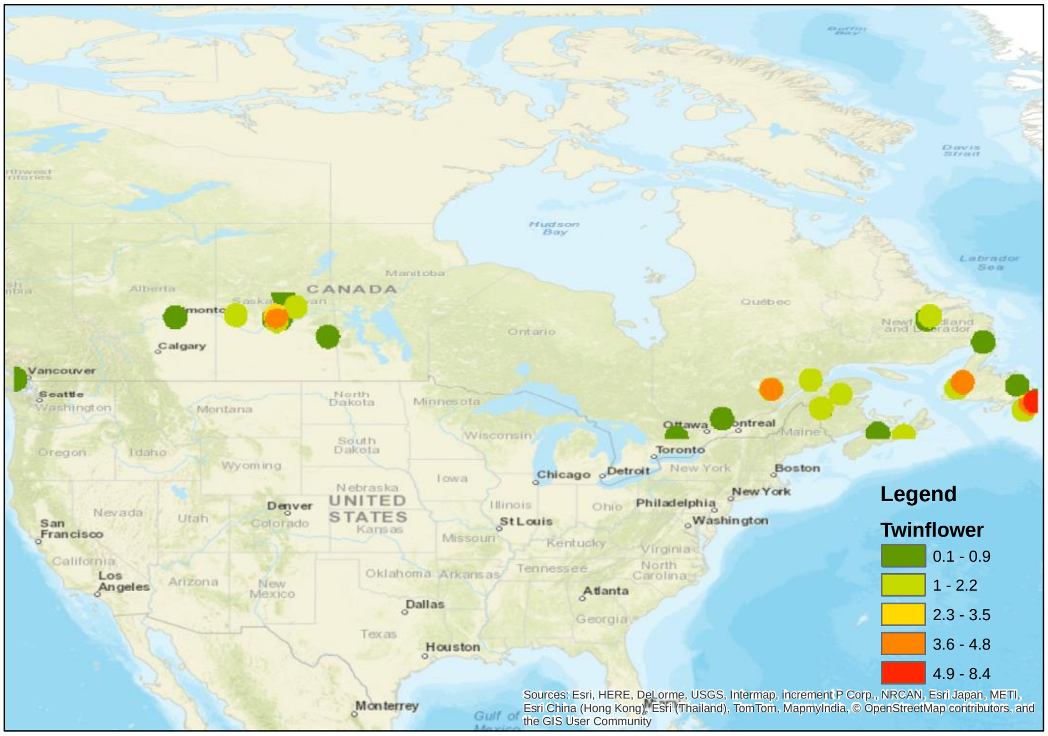

Figure 141: Twinflower entry density map.

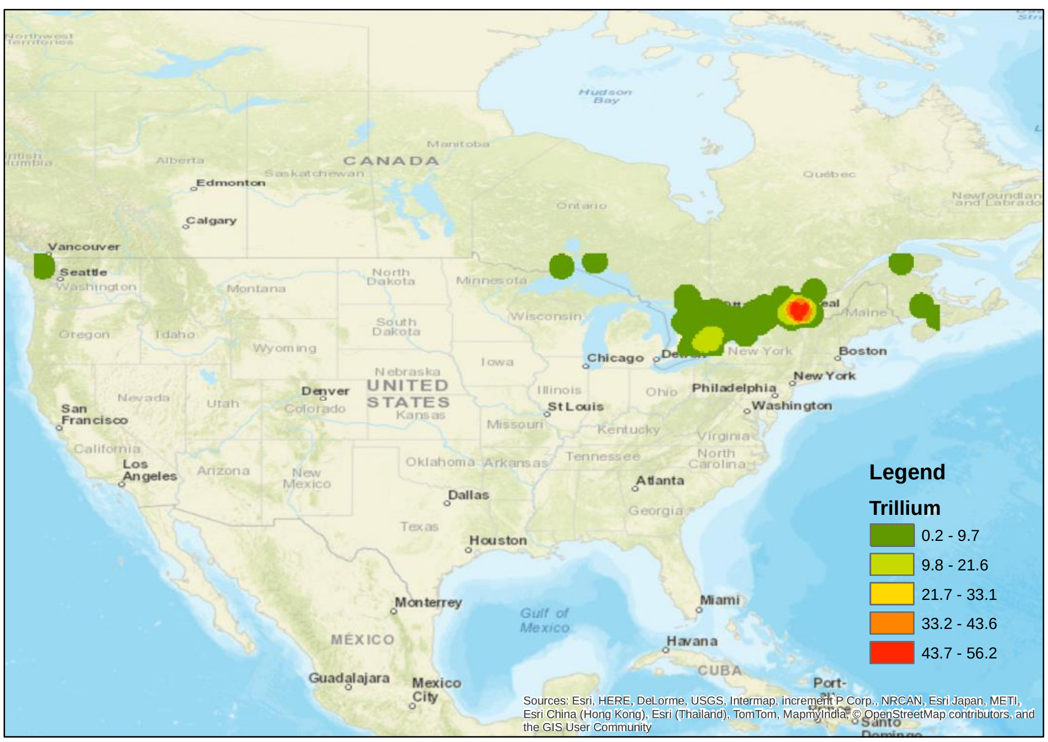

Figure 142: Entry density map for Trillium.

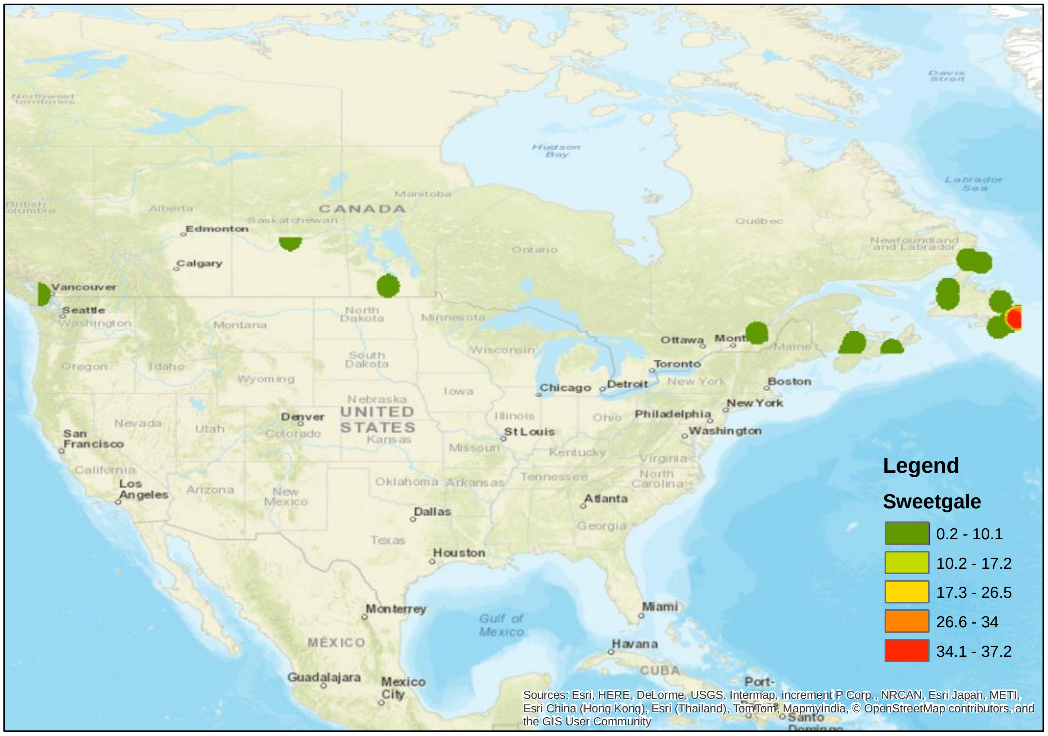

Figure 143: Map of entry density for Sweetgale.

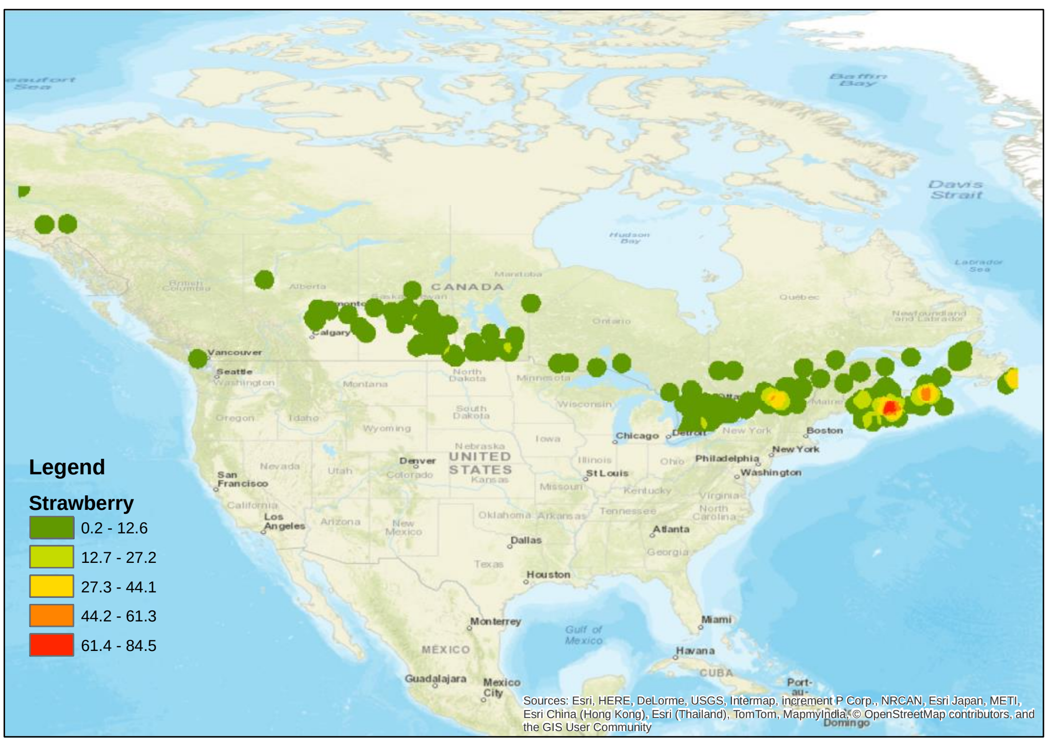

Figure 144: Strawberry entry density map.

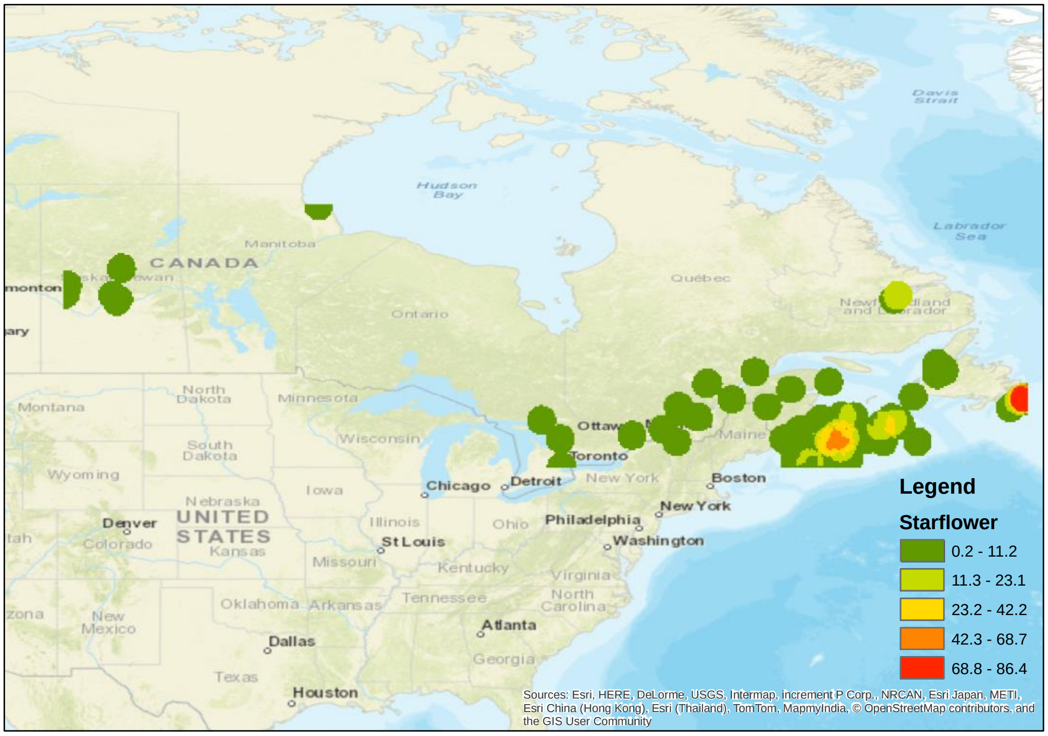

Figure 145: Starflower entry density map.

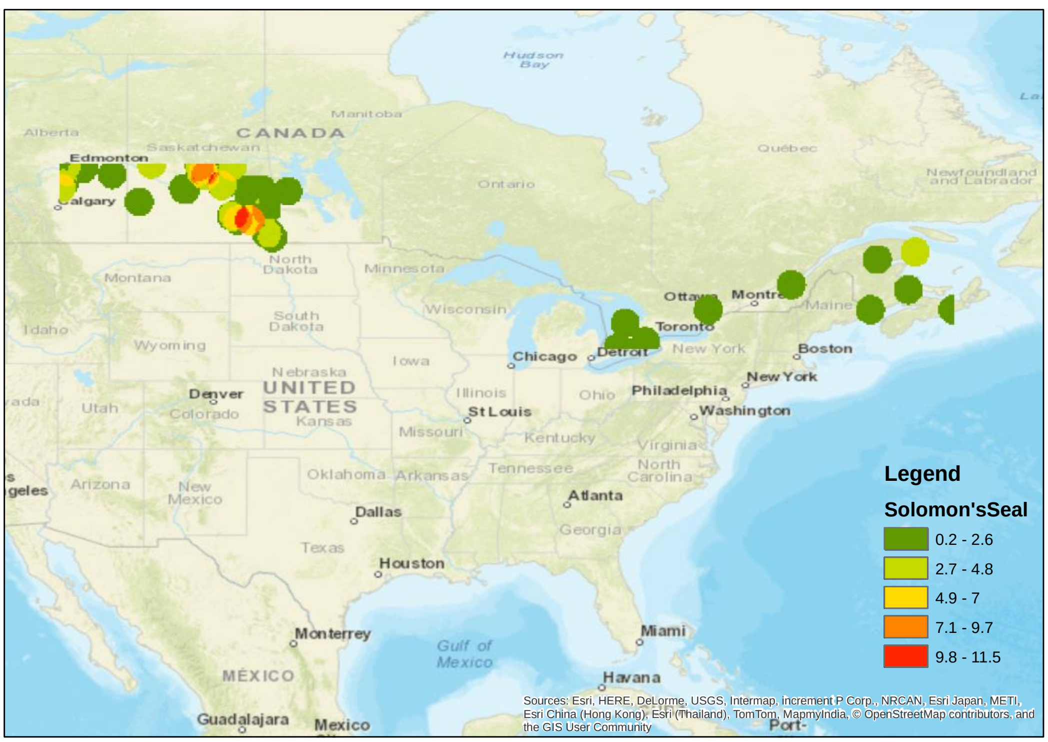

Figure 146: Entry density map for Solomon's Seal

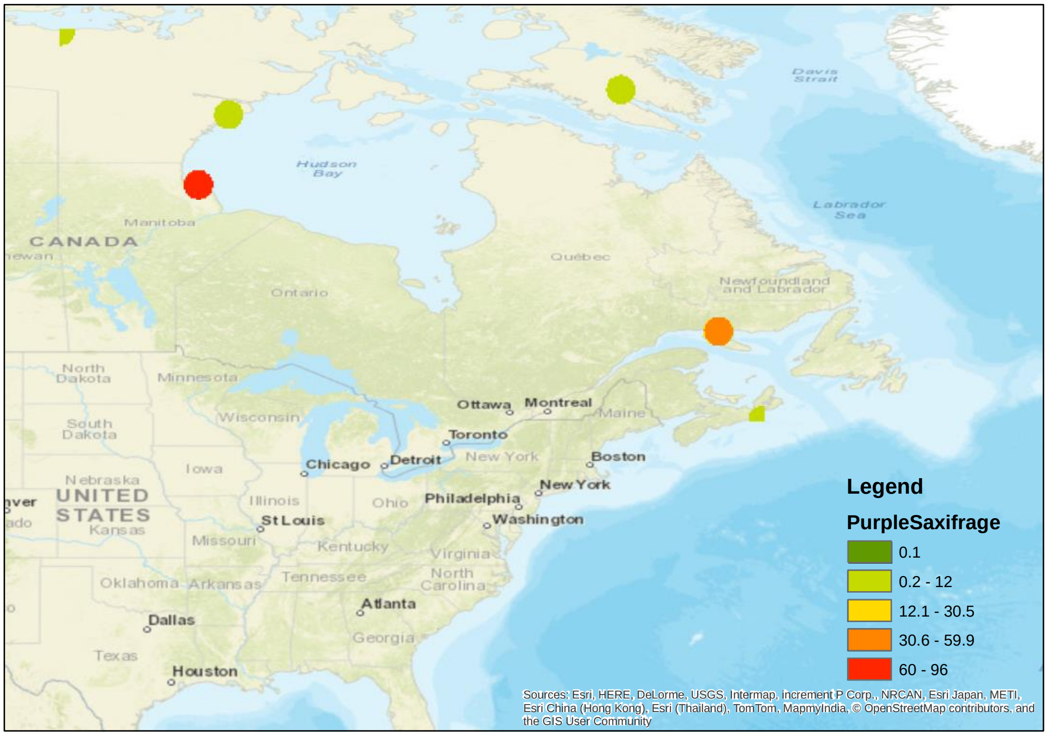

Figure 147: Purple saxifrage entry density map.

As we can see from this map there is a bit of a blip with this species in that there is a large concentration of entries in one single point in northeast Manitoba.

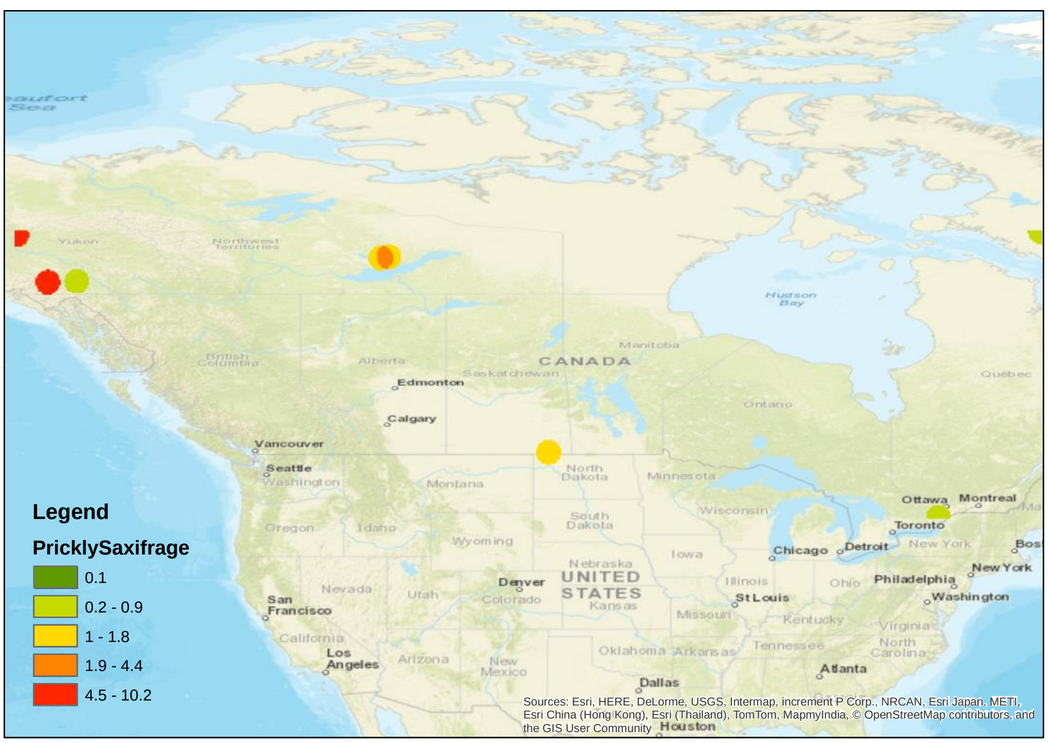

Figure 148: Prickly saxifrage entry density map.

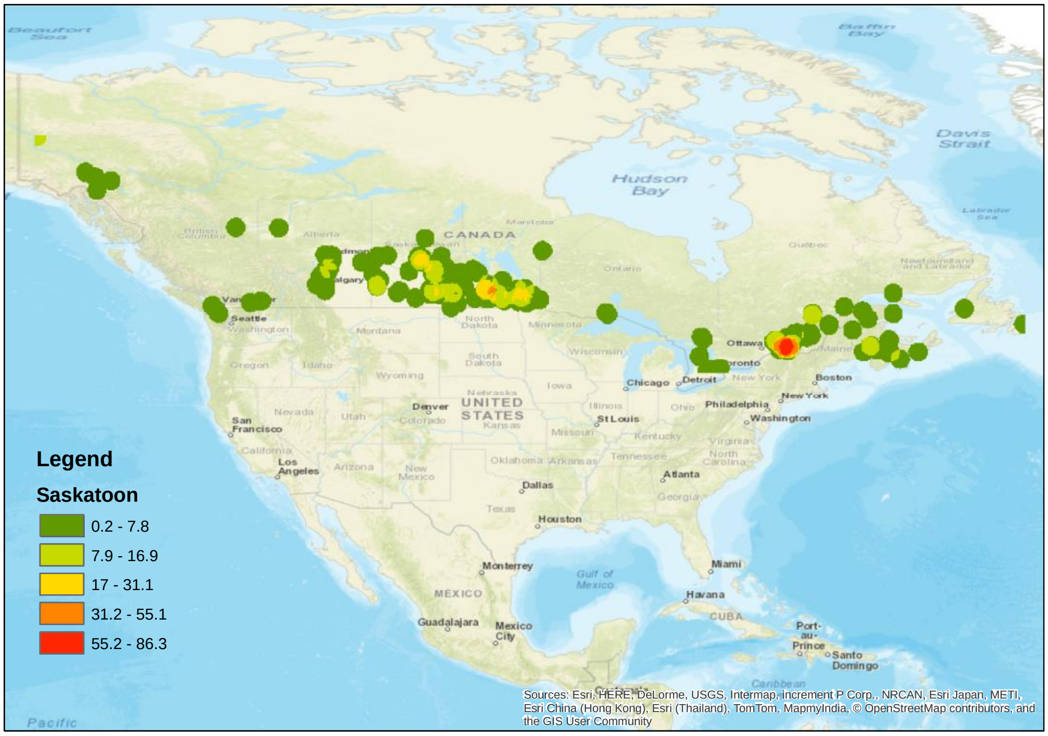

Figure 149: Entry density map for Saskatoon.

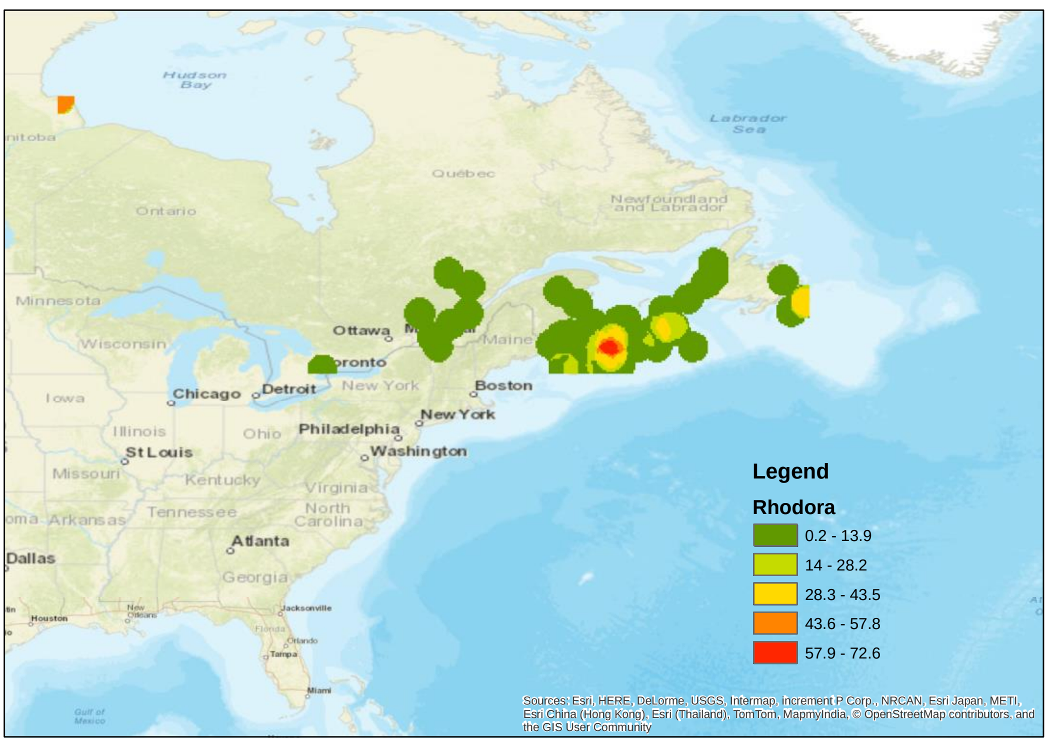

Figure 150: Rhodora entry density map.

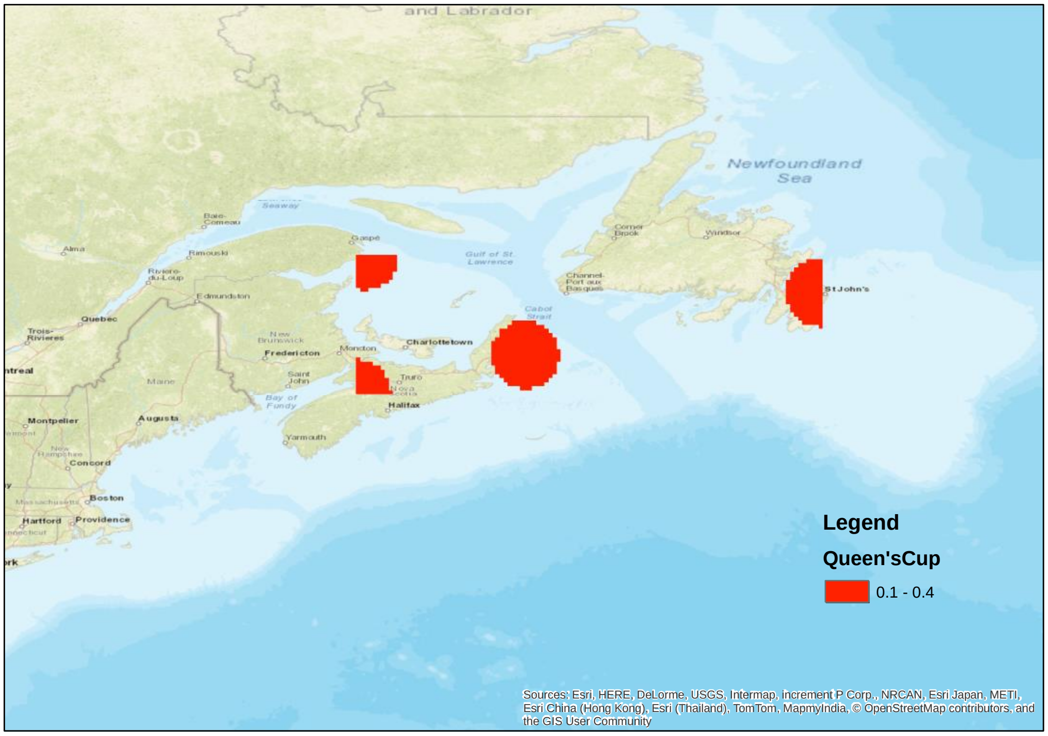

Figure 151: Queen's cup entry density map.

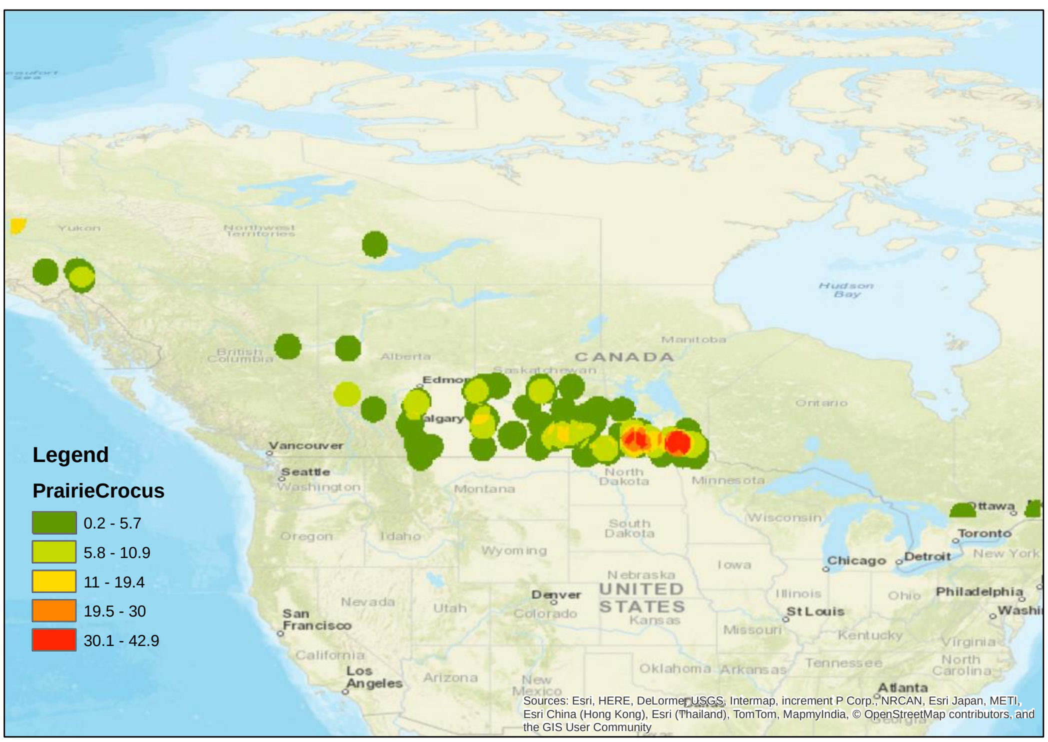

Figure 152: Entry density map for Prairie Crocus.

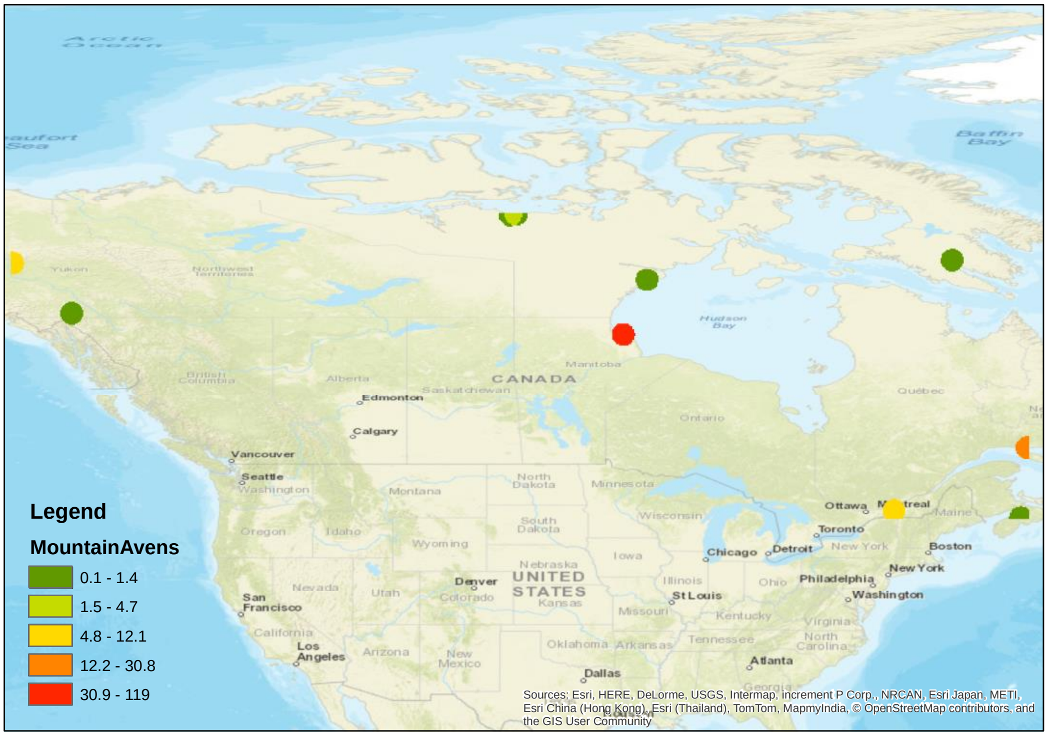

Figure 153: Mountain avens entry density map.

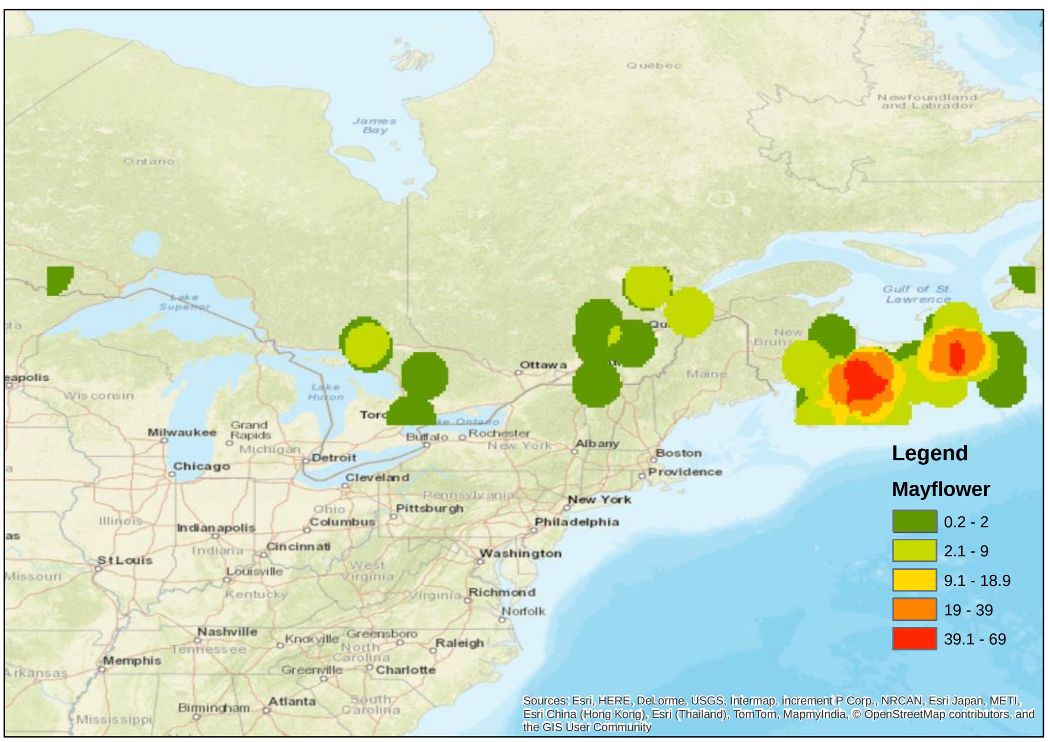

Figure 154: Mayflower entry density map.

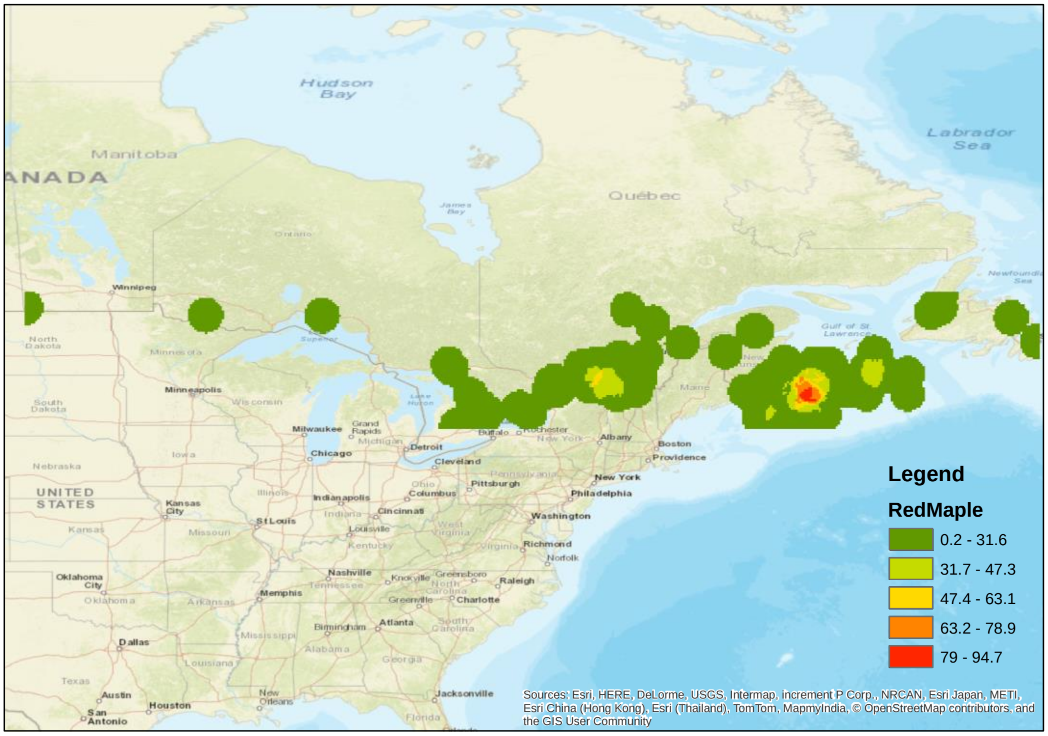

Figure 155: Red maple entry density map.

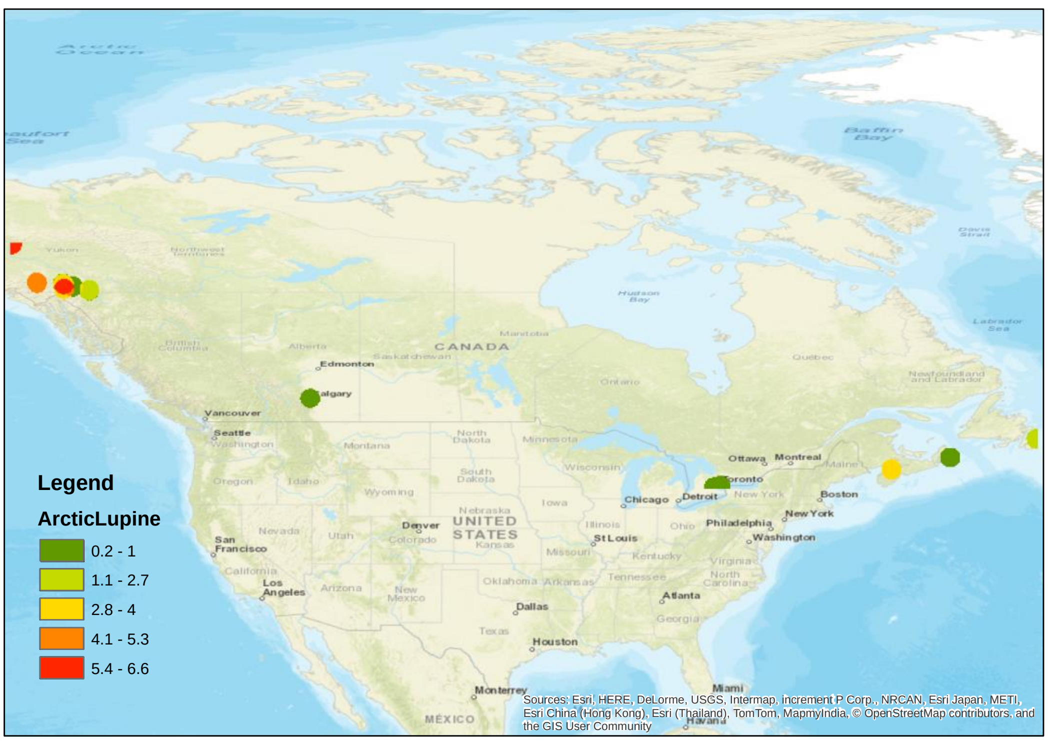

Figure 156: Arctic lupine entry density map.

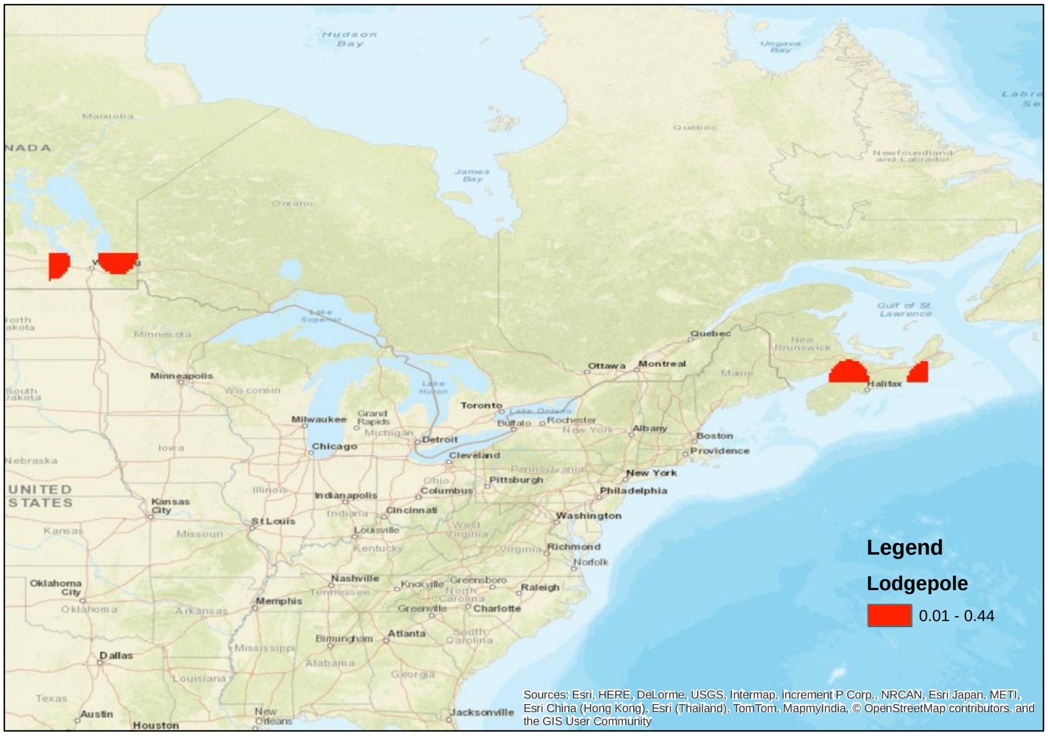

Figure 157: Lodgepole pine entry density map.

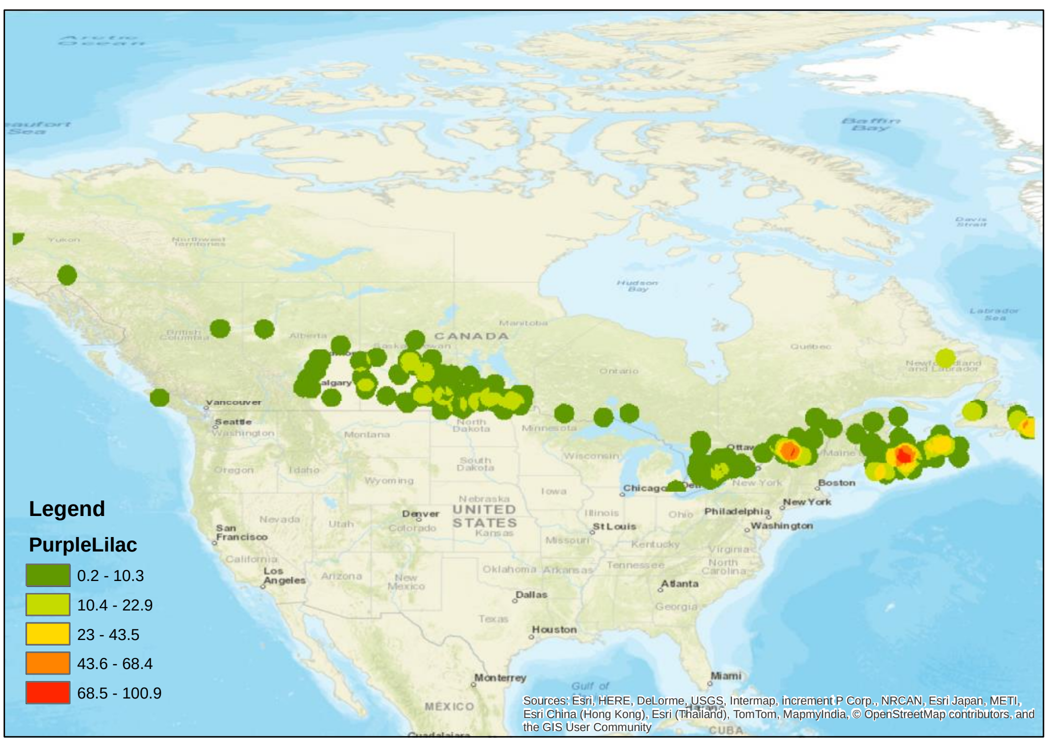

Figure 158: Entry density map of Purple lilacs.

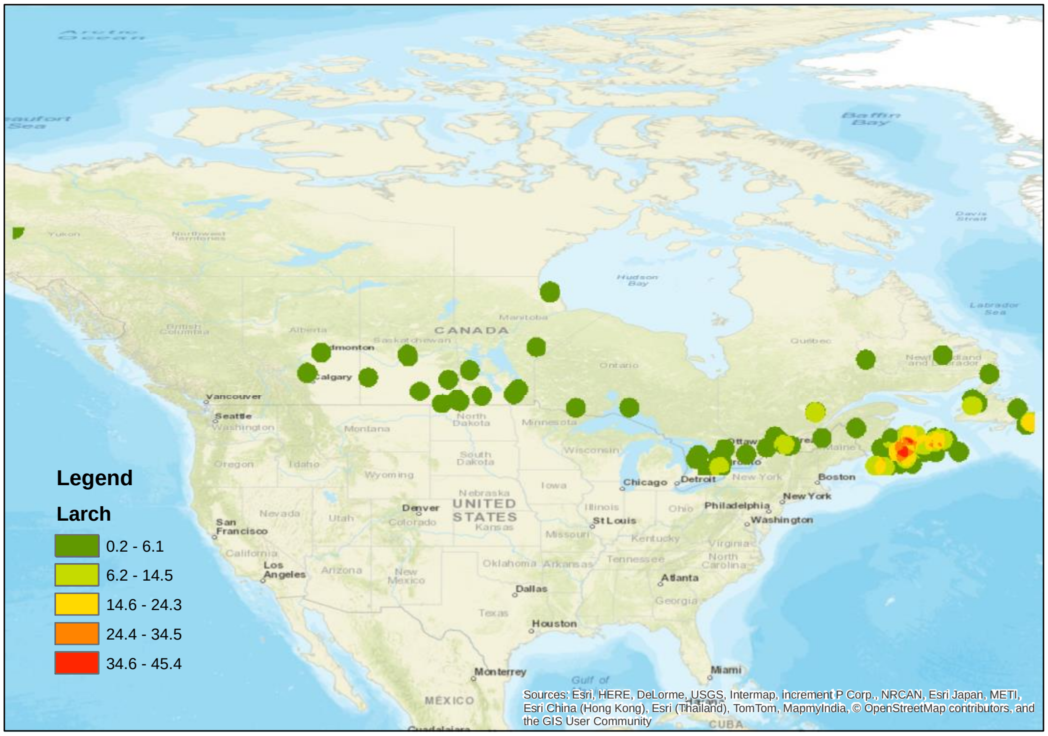

Figure 159: Larch entry density map.

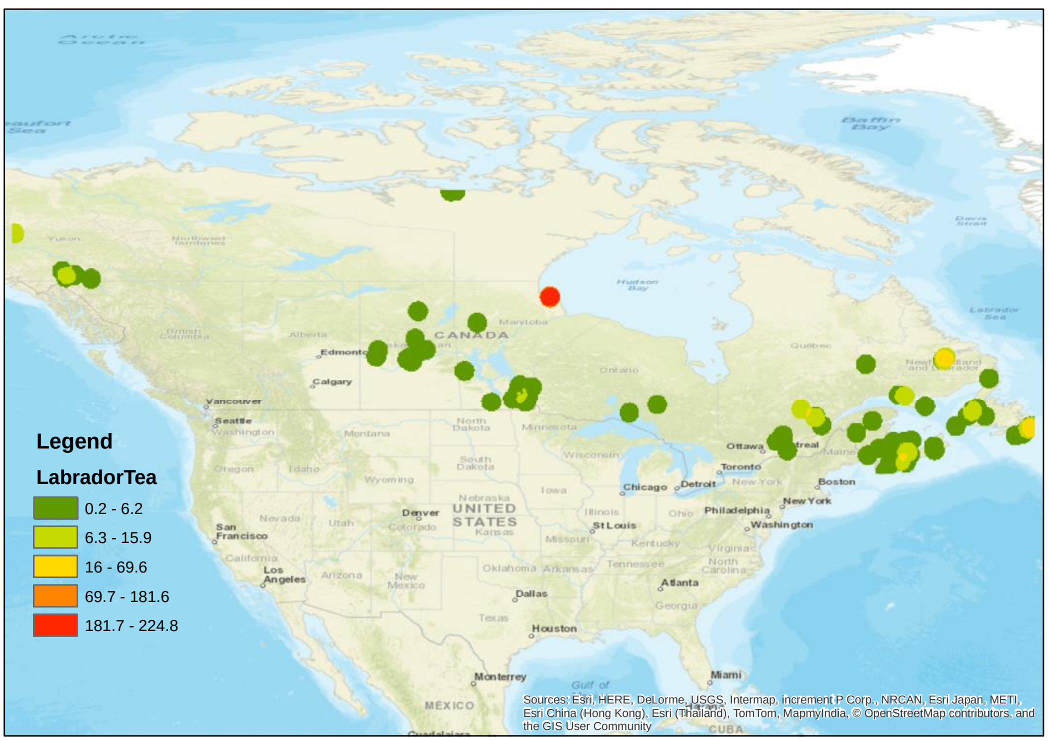

Figure 160: Labrador tea entry density map.

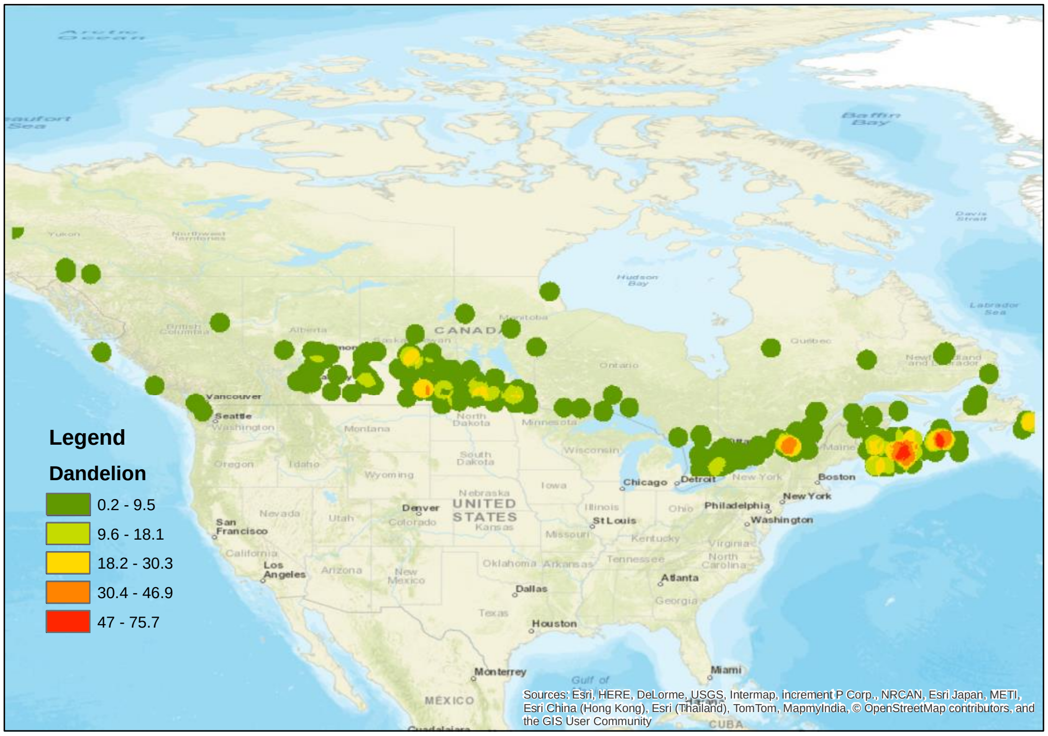

Figure 161: Entry density map for the Dandelion.

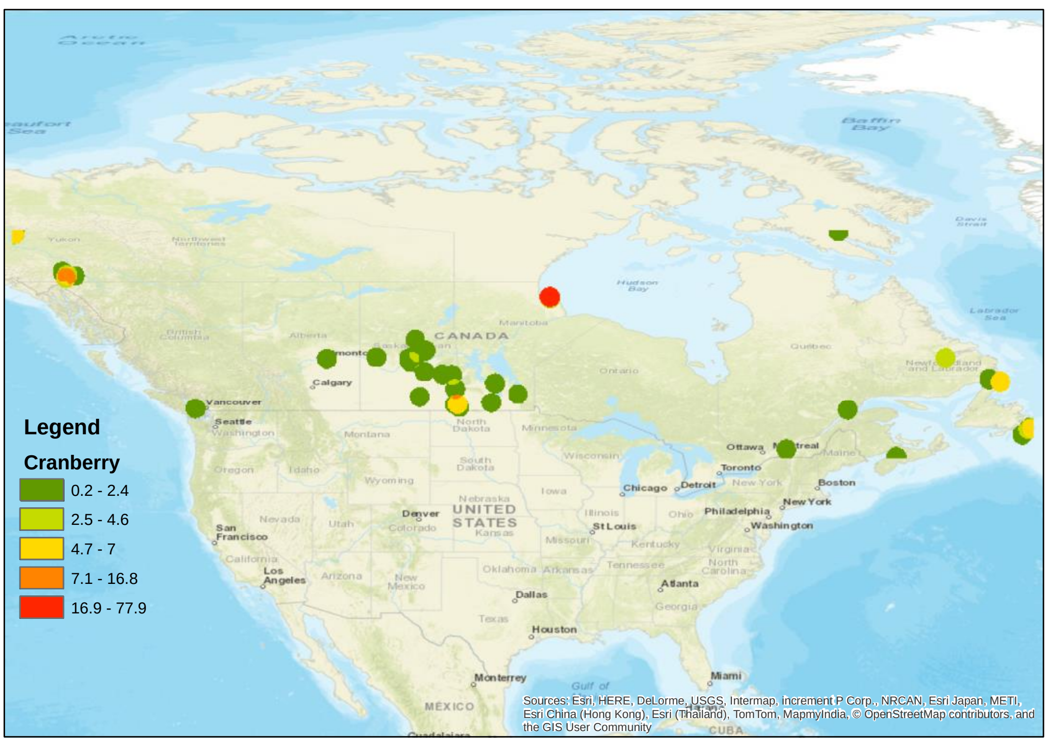

Figure 162: Cranberry entry density map.

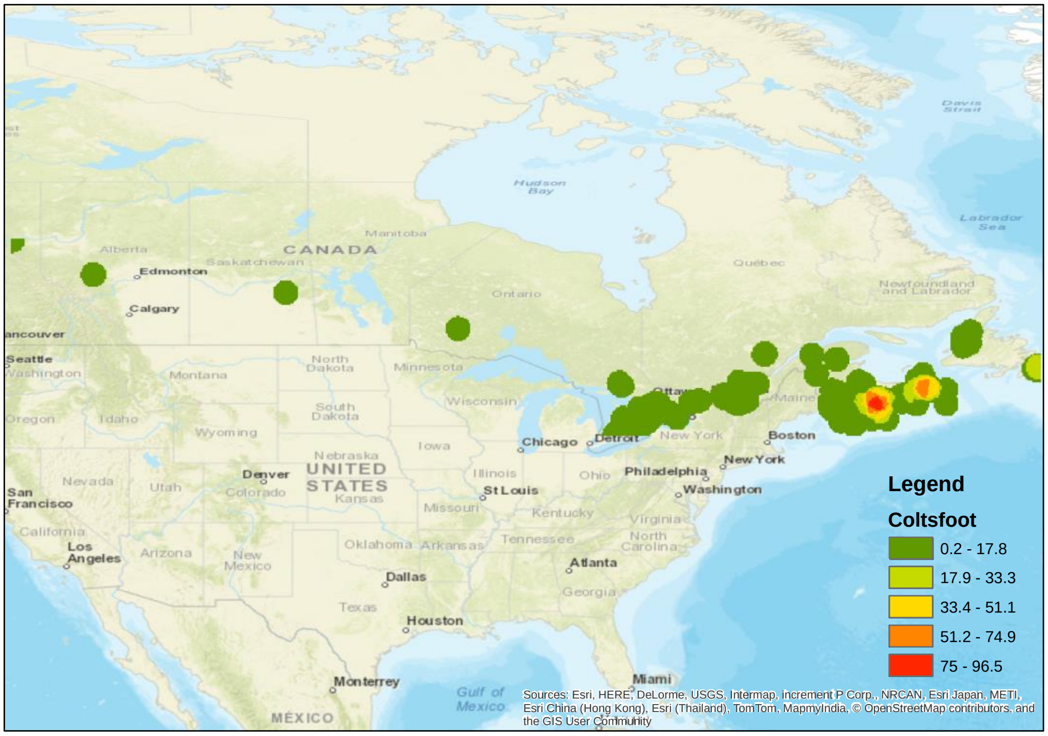

Figure 163: Coltsfoot entry density map.

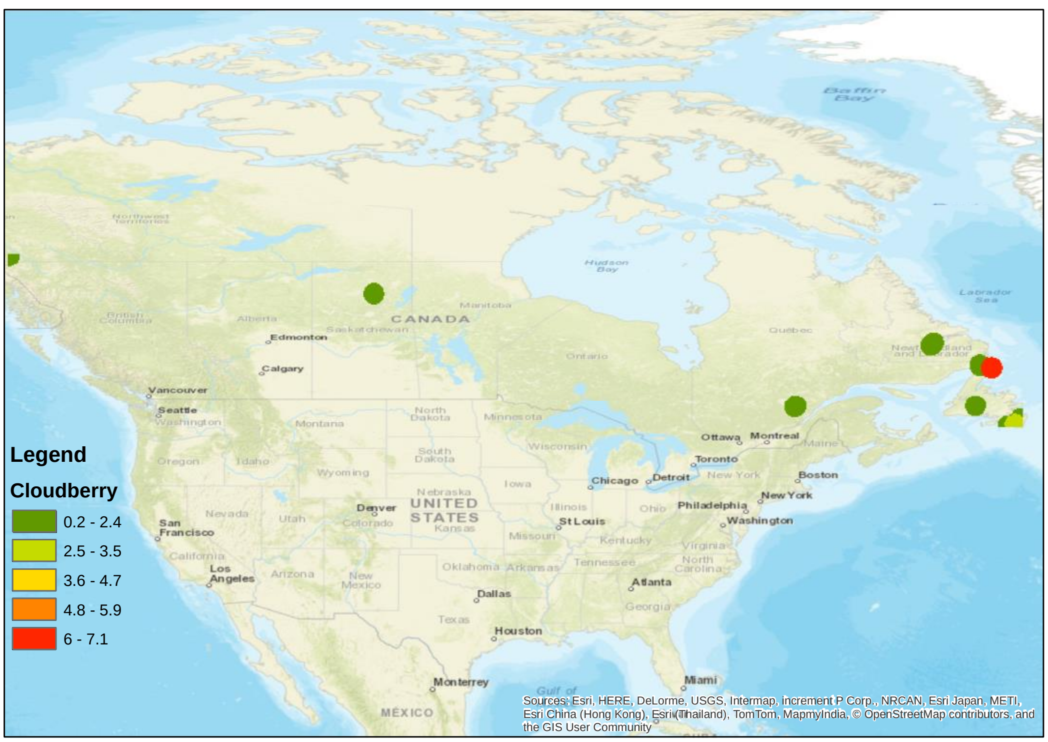

Figure 164: Entry density map for Cloudberry.

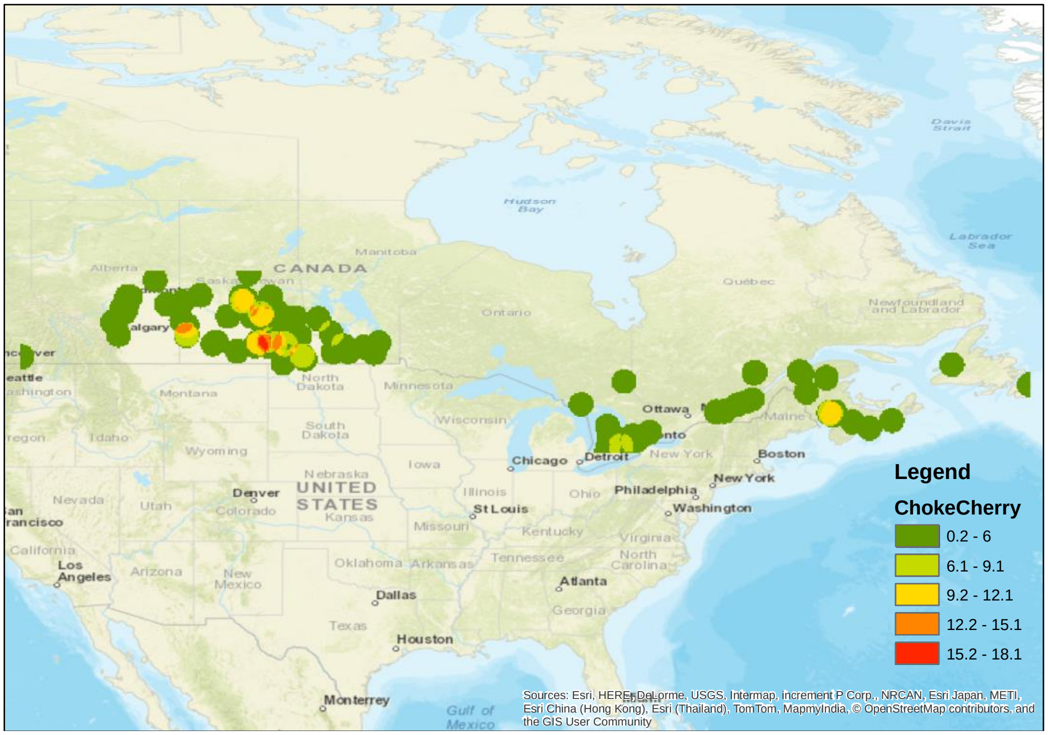

Figure 165: Choke cherry entry density map.

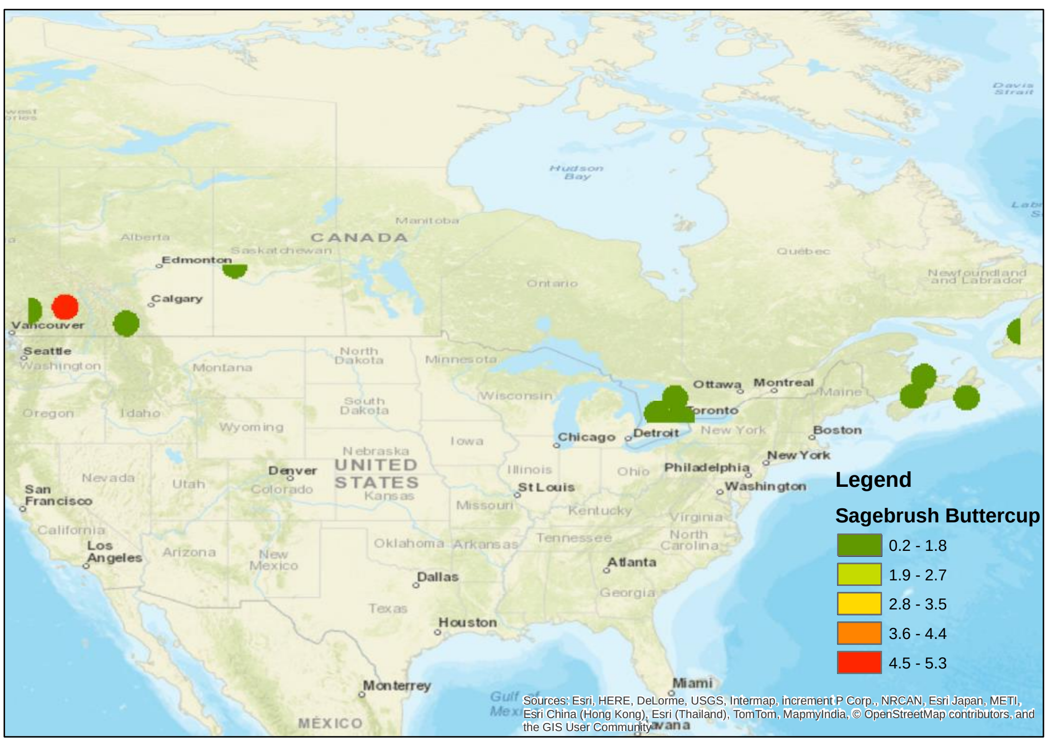

Figure 166: Sagebrush buttercup entry density map.

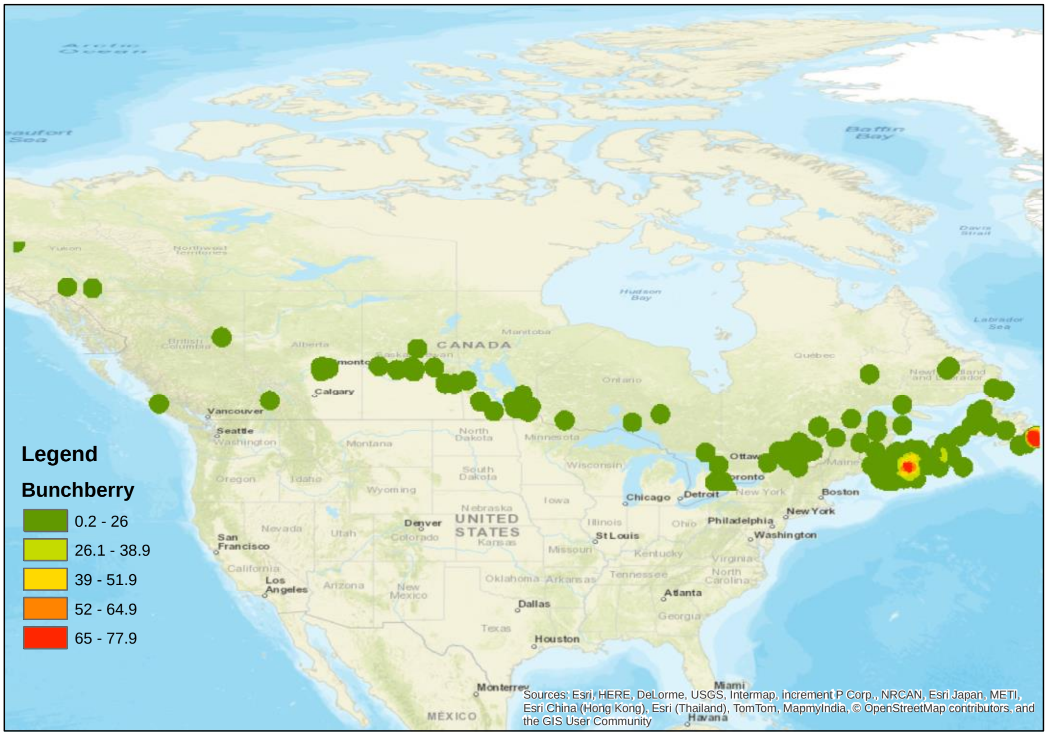

Figure 167: Entry density map for Bunchberry.

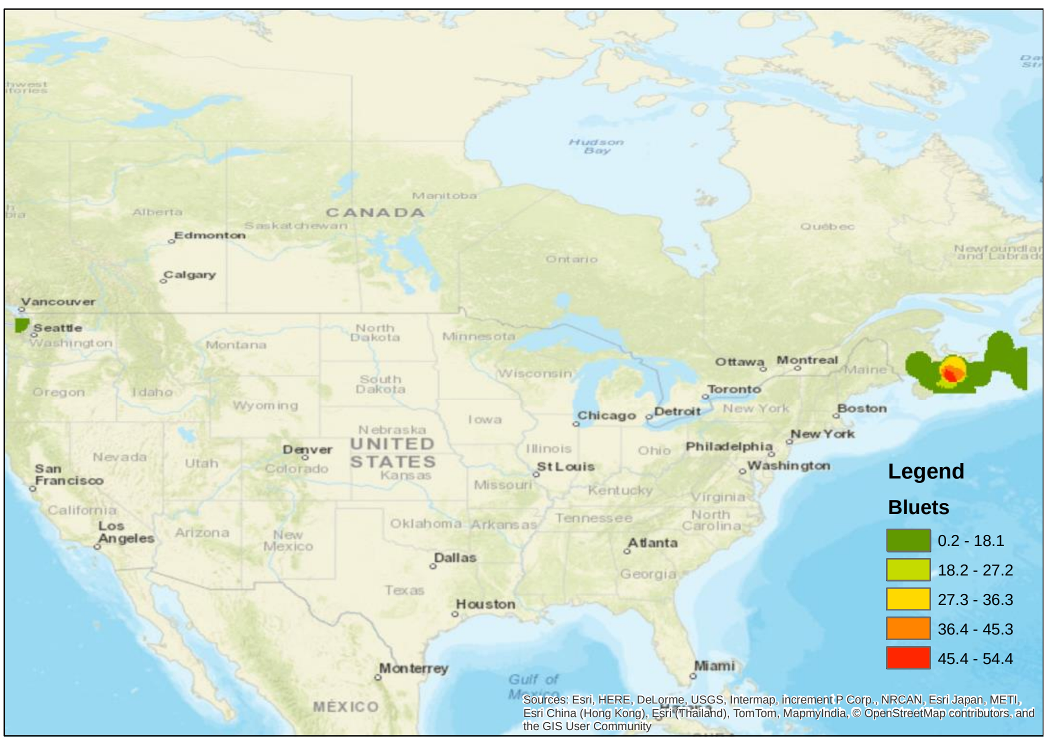

Figure 168: Bluets entry density map.

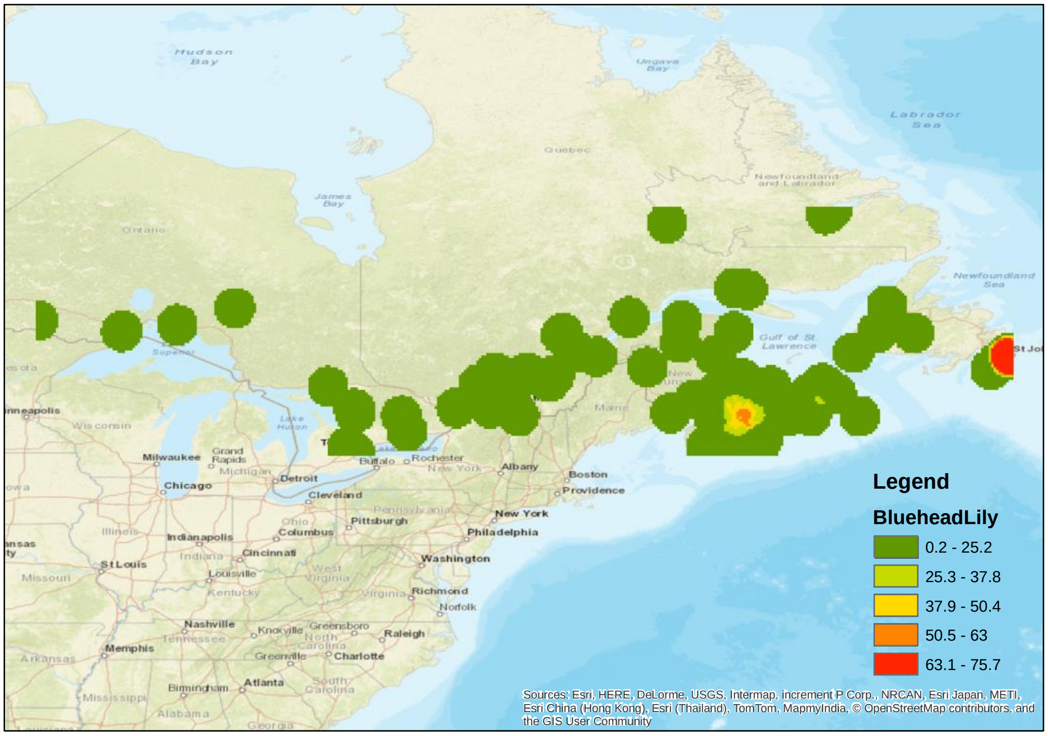

Figure 169: Bluehead lily entry density map.

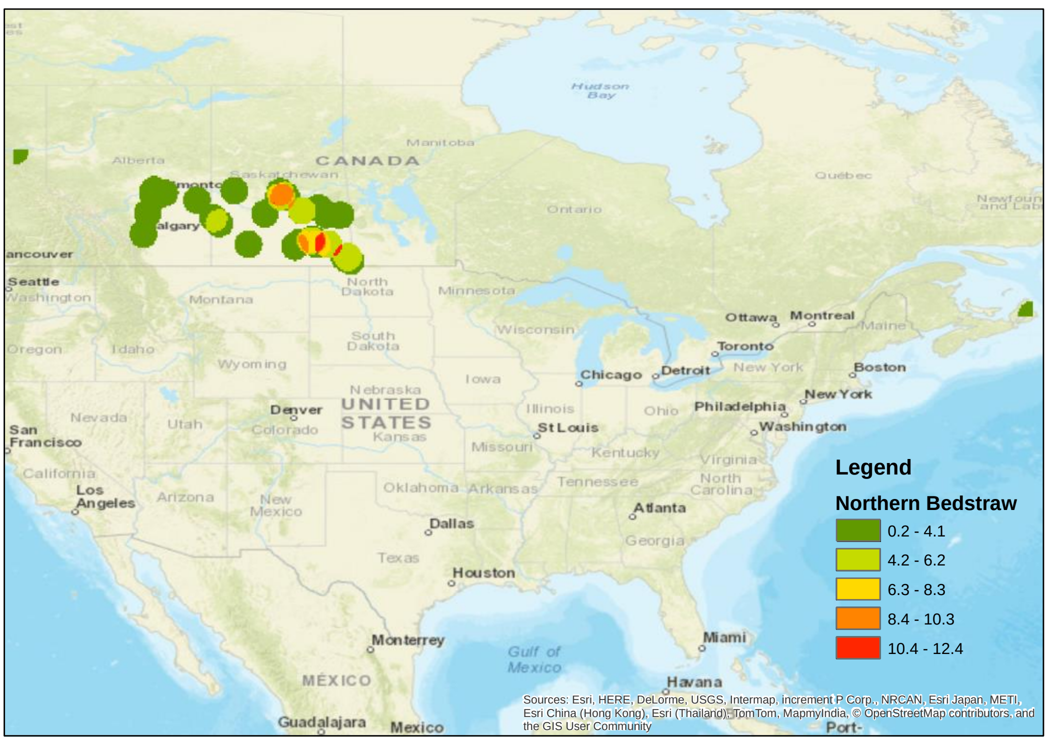

Figure 170: Entry density map for Northern bedstraw.

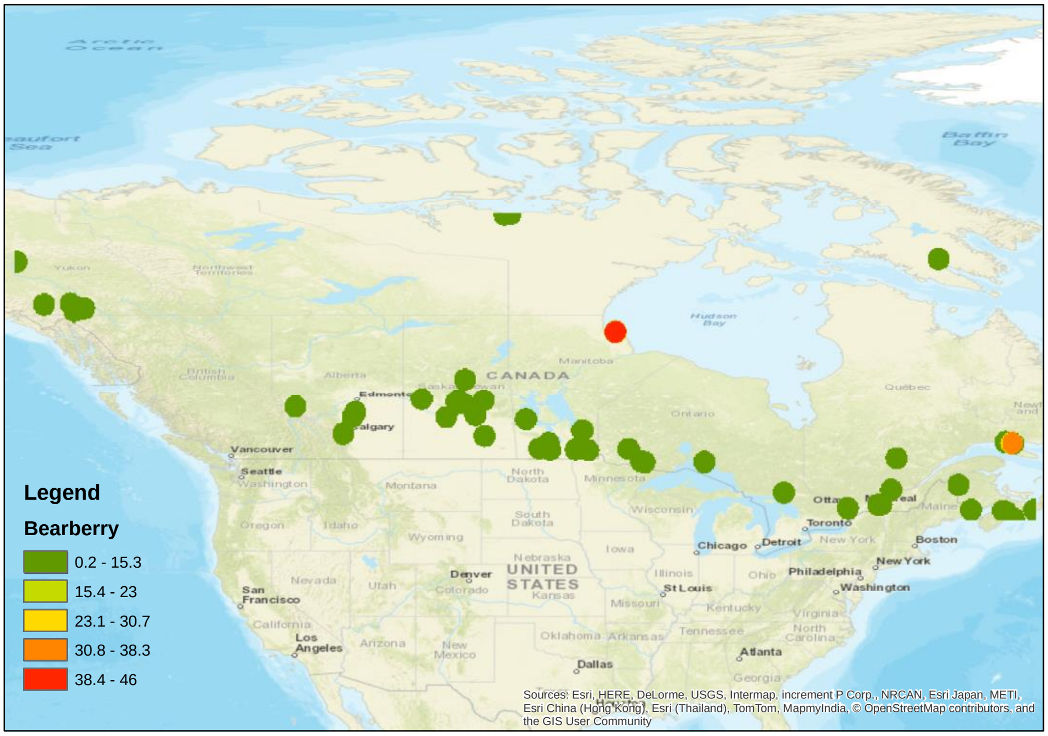

Figure 171: Bearberry entry density map.

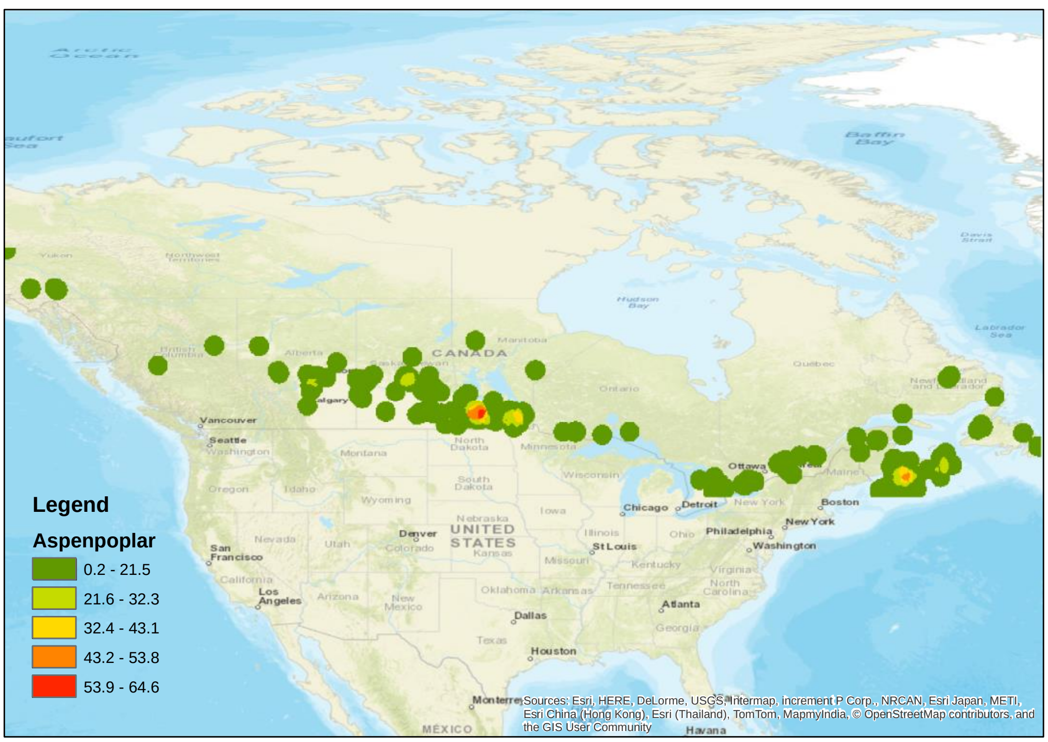

Figure 172: Aspen poplar entry density map.

Appendix B-4: Plantwatch observer analysis

We have conducted the observer analysis for Frogwatch already, and it indicated the general pattern wherein the majority of observers had 10 or fewer entries, but that the majority of entries came from the observers with contributions of more than 10 entries (anywhere between 11 and 1932 entries – with the extreme high end having a shockingly high percentage). By plotting the percentage of the total entry count for each "category" of observer (based on how many entries the observers in the category had), we end up with the following graph.

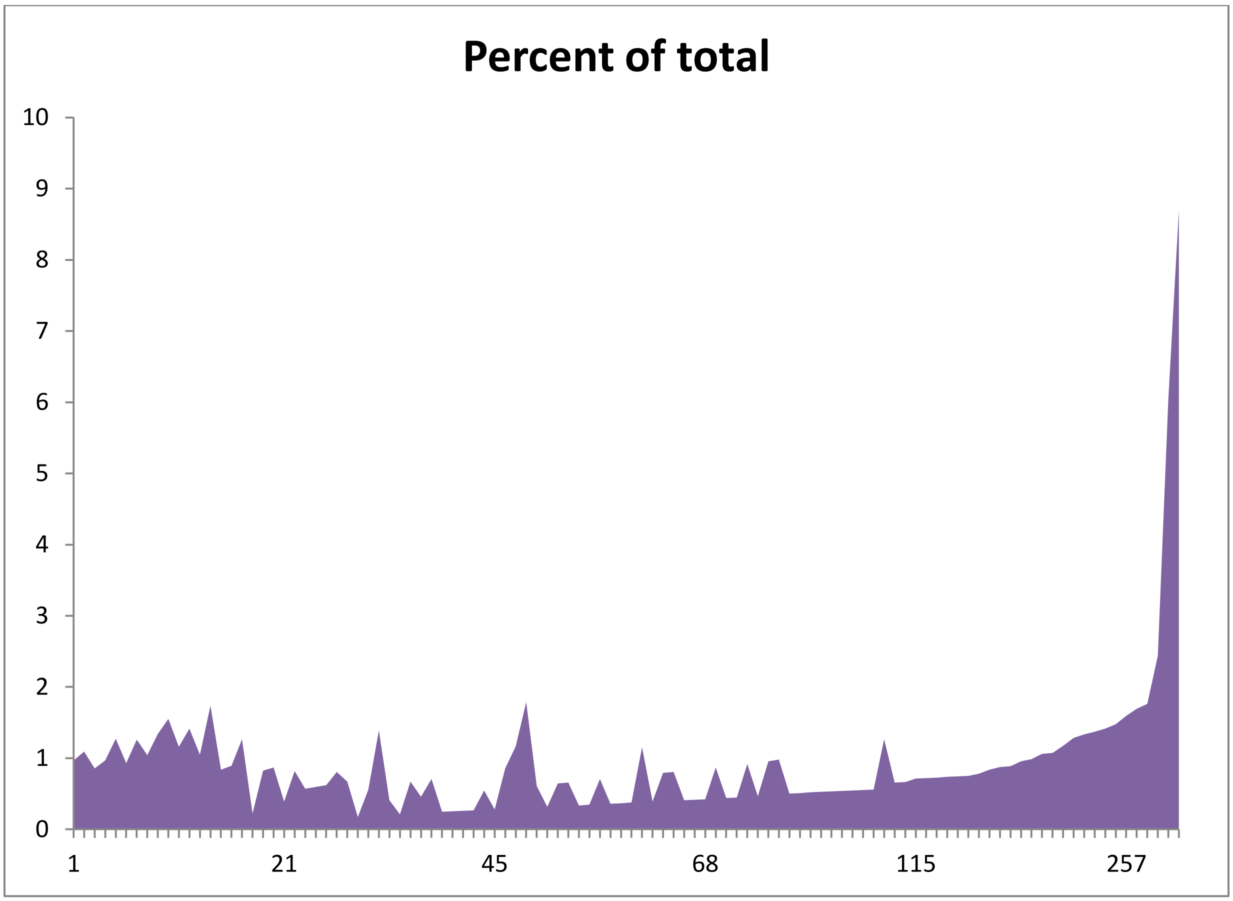

Figure 173: Percent of total entries originating from users with a given entry contribution.

As we can see from the graph, the trend observed for the FrogWatch contributors holds. Here we again see that the most prolific observers hold a disproportionate share of the entries. In fact, the 494 observers with 10 or fewer entries account for only 11,28% of all the entries, while the 262 with 11 or more entries accounted for 88,72% of the entries. What we do see as well, however, is that the casual observers around 65% of observers) were not in as crushing a majority as they were for Frogwatch (where that number was around 80%), which further reinforces the point that the dedicated observers are those to whom the program must reach out by priority.

Appendix C: Icewatch analysis

1.1 icewatch_observation

Table: icewatch_observation		
Column heading	**Data type**	**Reference table**
Pk_icewatch_observation	Integer	
Fk_location	Integer	location
Fk_observer	Integer	observer
Fk_icewatch_observation_code	Integer	Icewatch_observation_code
Create_date	Datetime	
Change_date	Datetime	
Visit_date	Datetime	
Air_temperature		
Water_temperature		
Comments		
Possible_outlier		
validated		
Audit_date		
image	Varchar(255)	
Fk_administrator_auditor		Missing

Pk_icewatch_observation: Number code assigned to each new observation as it is entered into the database.
Create/change_date: Refers to when the entry was assigned a pk_icewatch_observation value and when the details have been edited where relevant.
Visit_date: Refers to the date the observation was made as entered by the contributor.
Air/water_temperature: Entered by the contributor, defines the air or water temperatures at the time of the visit. This is not filled for most entries (only some 1500 of the air temperature, and 117 for water, out of over 10 000).
Comments: Personalised comment, if any.
Possible_outlier: Indicates if the entry is an outlier.
Validated: Indicates whether the entry has been validated. Most entries are "NULL".
Audit_date: Defunct column.
Image: Image file, if any.
Fk_administrator_auditor: All the entries are NULL. Defunct column

1.2 location

Table: location		
Column heading	**Data type**	**Reference table**
Pk_location	Integer – auto-increment	
Fk_observer	Integer	Observer
Create_date	datetime	

Change_date	datetime	
Location_name	Varchar(255)	
Observer_location_code	Varchar(255)	
Nearest_named_place	Varchar(255)	
Nnp_direction	Varchar(2)	
Nnp_distance	Integer	
latitude	Decimal(18,12)	
longitude	Decimal(18,12)	
Coordinate_source_code	Varchar(2)	
Coordinate_accuracy_code	Varchar(3)	
Location_description	Varchar(255)	
comments	Varchar(255)	
Location_type_code	Char(1)	
elevation	Varchar(20)	
province	Varchar(50)	

Pk_location: Numerical code assigned to each new location, as described in this chapter's introduction.
Create/change_date: Date at which the location was entered into the database and assigned its code.
Location_name: Name of the location as provided by the observer. This can be a general location such as the nearest city or a very specific location such as "[person]'s backyard".
Observer_location_code: Defunct column – none of the entries have non-null values. This is because observers were formerly associated with a single location in the past, while now observers are considered wholly independent from the locations where their observations are made.
Nearest_named_place: This column allows the observer to make locating where the observation was made in the event it is some distance from the "nearest named" place. Usually used in a complimentary manner with the location_name in the event it is too specific/local.
Nnp_direction: Direction to the nnp (nearest named place), using cardinal points.
Nnp_distance: Distance to the nnp.
Latitude/longitude: Coordinates provided by the observer.
Coordinate_source_code: Source of the coordinate data, as provided by the observer. It must be noted that only 191 of the locations have a value entered into this column, out of over 24 500 locations, meaning the coordinates cannot be verified with certainty.
Coordinate_accuracy_code: All values are NULL. Defunct column.
Location_description: Observers can describe the location in this column to provide some additional information.
Comments: The observer can add information here that isn't included in the other columns. All values are NULL.
Location_type_code: 1-digit letter code assigned to the location. Only provided in 100 locations out of the over 24 500, and the data itself is not clearly identified. Defunct column.
Elevation:
Province: Name of the province (or territory) where the observation was made.

1.3 observer

Table: observer		
Column heading	**Data type**	**Reference table**
Pk_observer	Integer – auto-increment	
Fk_partner	Integer	Partner
Fk_age_range	Integer	Age_range

Observer_code	Varchar(255)	
Observer_language	Varchar(1)	
Create_date	Datetime	
Change_date	Datetime	
Observer_first_name	Varchar(255)	
Observer_last_name	Varchar(255)	
Mailing_address_street	Varchar(255)	
Mailing_address_city	Varchar(255)	
Mailing_address_province_code	Varchar(2)	
Mailing_address_country_code	Varchar(3)	
Mailing_address_postal_code	Varchar(255)	
Telephone_number	Varchar(255)	
Email_address	Varchar(128)	
Affiliation	Varchar(128)	
Comments	Varchar(254)	
Send_newsletter	Varchar(10)	
Password	Varchar(512)	
Password_type	Varchar(50)	
Password_temporary	Varchar(16)	
Email_temp	Varchar(255)	
Temp_password_sent	Varchar(255)	
Old_observer_code	Varchar(255)	
Email_validated	Tinyint(4)	

Pk_observer: This column is a numbered identifier for each contributor/observer. The increments are 1 per observer, meaning every observer whose information does not match a pre-existing one is assigned a number 1 higher than the previous one.
Observer_code: Largely analogous to the pk_observer column on a one-for-one basis, except the starting point is much higher (1111111 as opposed to 21).
Observer_language: Possible entries are E and F, for English and French. NULL entries occur when no language is specified.
Create/change_date: Refers to the date the observer was entered into the database, or edited.
Observer_first_name, Observer_last_name: Where provided, identifies the observer's full name via these two columns.
Mailing_address_/city/province_code/country_code/postal_code: Provides the respective mailing address for each category where provided.
Telephone_number, email_address: Where provided, contains the observer's contact information.
Affilitation: The observer's affiliation to an organisation, which can be a private company, corporation or non-profit organisation, among others.
Comments:
Send_newsletter: Notes whether the observer wishes to receive newsletters published on NatureWatch.
Password, Password_type, Password_temporary: The observer's password to sign into the website, for access to the data and to enter observations. The temporary password is the one sent automatically to the observer upon registration, where the "password" variable is null unless the observer changed it to a personalised password.
Email_temp: Temporary password automatically sent to the observer upon registration.

Temp_password_sent: This column informs on whether the observer received a confirmation email for his/her registration as an observer for NatureWatch.
Old_observer_code: The database was altered over the years, including a streamlining of the process of assigning numbers to observers. This column contains the numbers for the observers prior to the update, and so more recently added observers do not have a number in this column.
Email_validated: Denotes if the observer received and acknowledged the confirmation email for the creation of his or her account on the NatureWatch website.

1.4 icewatch_observation_code

Table: icewatch_observation_code		
Column heading	**Data type**	**Reference table**
Pk_icewatch_observation_code	Integer	
English_name	Varchar(255)	
French_name	Varchar(255)	
English_description	Varchar(255)	
French_description	Varchar(255)	

Pk_icewatch_observation_code: Number code assigned to a particular observation circumstance for ice conditions.
English/French_name: Name given to the ice conditions in English and French.
English/French_description: Short description of the conditions, in English and French.

1.5 Partner

Table: partner		
Column heading	**Data type**	**Reference table**
Pk_partner	Integer	
Create_date	Datetime	
Change_date	Datetime	
English_name	Varchar(255)	
French_name	Varchar(255)	
Mailing_address_street	Varchar(255)	
Mailing_address_city	Varchar(64)	
Mailing_address_province_code	Varchar(2)	
Mailing_address_country_code	Varchar(3)	
Mailing_address_postal_code	Varchar(6)	
Telephone_number	Varchar(10)	
Fax_number	Varchar(10)	
Email_address	Varchar(128)	
Contact_name	Varchar(128)	
Comments	Varchar(255)	

Pk_partner: Number code assigned to the program partner.
Create/change_date: Date when the pk_partner entry was entered or its details changed.
English/French_name: English and French names of the partner organisation.
Mailing_address_street/city/province_code/country_code/postal_code: Coordinates of the partner.
Telephone/fax_number, email_address: Contact information.
Comments:

It must be noted that there are only 4 entries in this table, none of which contain any information besides the pk_partner and create/change date. The table as a whole is thus defunct.

1.6 age_range

Table: age_range		
Column heading	**Data type**	**Reference table**
Pk_age_range	Integer	
Range_text	Varchar(50)	

Pk_age_range: Assigns an increment code to the age of the observer.
Range_text: The values (high and low) delimiting the above increments.

Appendix D: Wormwatch tables

1.1 wormwatch_observation

Table: wormwatch_location		
Column heading	**Data type**	**Reference table**
Pk_wormwatch_observation	Integer – auto-increment	
Fk_location	integer	location
Fk_observer	integer	observer
Fk_visit	integer	visit
Sampling_technique	Varchar(50)	
Hole_length	Varchar(12)	
Hole_width	Varchar(12)	
Hole_depth	Varchar(12)	
Adult_species	integer	
Number_observed	integer	
Total_number_adults	integer	
Total_number_juveniles	integer	
Total_number_cocoons	integer	
aestivating		
comments		
Fk_soil_type		soil_type
Fk_soil_colour		soil_colour
Create_date		
Change_date		
Visit_date		
Fk_worm_species		worm_species
image		

Pk_wormwatch_observation: Numerical code assigned to each new location, as described in this chapter's introduction.
Sampling_technique: Describes the basic technique used to sample the worms, the three options being "Modified hand-sort", "quantified hand-sort" and "flip and strip".
Hole_length/width/depth: Includes measurements in centimeters for the parameters of the hole dug to sample worms from.
Adult_species: Contains the number code for the worm species – note that this column is redundant as it contains identical information to the fk_worm_species column.
Number_observed: Describes the total number of worms of the given species found in the sample.
Total_number_adults/juveniles/cocoons: Contains the number of each worm of the identified species at the stage of development specified.
Aestivating: Indicates whether aestivating specimens are present or absent.
Comments: Includes personal comments about the sample or circumstances of the observation.
Create/Change/visit_date: Date at which the entry was created in the database and modified. The visit date is entered by the observer and indicates the actual date at which the sample was taken and analysed.
Image: Image file, if any.

1.2 location

Table: location		
Column heading	**Data type**	**Reference table**
Pk_location	Integer – auto-increment	
Fk_observer	Integer	Observer
Create_date	datetime	
Change_date	datetime	
Location_name	Varchar(255)	
Observer_location_code	Varchar(255)	
Nearest_named_place	Varchar(255)	
Nnp_direction	Varchar(2)	
Nnp_distance	Integer	
latitude	Decimal(18,12)	
longitude	Decimal(18,12)	
Coordinate_source_code	Varchar(2)	
Coordinate_accuracy_code	Varchar(3)	
Location_description	Varchar(255)	
comments	Varchar(255)	
Location_type_code	Char(1)	
elevation	Varchar(20)	
province	Varchar(50)	

Pk_location: Numerical code assigned to each new location, as described in this chapter's introduction.
Create/change_date: Date at which the location was entered into the database and assigned its code.
Location_name: Name of the location as provided by the observer. This can be a general location such as the nearest city or a very specific location such as "[person]'s backyard".
Observer_location_code: Defunct column – none of the entries have non-null values. This is because observers were formerly associated with a single location in the past, while now observers are considered wholly independent from the locations where their observations are made.
Nearest_named_place: This column allows the observer to make locating where the observation was made in the event it is some distance from the "nearest named" place. Usually used in a complimentary manner with the location_name in the event it is too specific/local.
Nnp_direction: Direction to the nnp (nearest named place), using cardinal points.
Nnp_distance: Distance to the nnp.
Latitude/longitude: Coordinates provided by the observer.
Coordinate_source_code: Source of the coordinate data, as provided by the observer. It must be noted that only 191 of the locations have a value entered into this column, out of over 24 500 locations, meaning the coordinates cannot be verified with certainty.
Coordinate_accuracy_code: All values are NULL. Defunct column.
Location_description: Observers can describe the location in this column to provide some additional information.
Comments: The observer can add information here that isn't included in the other columns. All values are NULL.
Location_type_code: 1-digit letter code assigned to the location. Only provided in 100 locations out of the over 24 500, and the data itself is not clearly identified. Defunct column.
Elevation: Elevation data where provided – most entries do not have this value.
Province: Name of the province (or territory) where the observation was made.

1.3 observer

Table: observer		
Column heading	**Data type**	**Reference table**
Pk_observer	Integer – auto-increment	
Fk_partner	Integer	Partner
Fk_age_range	Integer	Age_range
Observer_code	Varchar(255)	
Observer_language	Varchar(1)	
Create_date	Datetime	
Change_date	Datetime	
Observer_first_name	Varchar(255)	
Observer_last_name	Varchar(255)	
Mailing_address_street	Varchar(255)	
Mailing_address_city	Varchar(255)	
Mailing_address_province_code	Varchar(2)	
Mailing_address_country_code	Varchar(3)	
Mailing_address_postal_code	Varchar(255)	
Telephone_number	Varchar(255)	
Email_address	Varchar(128)	
Affiliation	Varchar(128)	
Comments	Varchar(254)	
Send_newsletter	Varchar(10)	
Password	Varchar(512)	
Password_type	Varchar(50)	
Password_temporary	Varchar(16)	
Email_temp	Varchar(255)	
Temp_password_sent	Varchar(255)	
Old_observer_code	Varchar(255)	
Email_validated	Tinyint(4)	

Pk_observer: This column is a numbered identifier for each contributor/observer. The increments are 1 per observer, meaning every observer whose information does not match a pre-existing one is assigned a number 1 higher than the previous one.
Observer_code: Largely analogous to the pk_observer column on a one-for-one basis, except the starting point is much higher (1111111 as opposed to 21).
Observer_language: Possible entries are E and F, for English and French. NULL entries occur when no language is specified.
Create/change_date: Refers to the date the observer was entered into the database, or edited.
Observer_first_name, Observer_last_name: Where provided, identifies the observer's full name via these two columns.
Mailing_address_/city/province_code/country_code/postal_code: Provides the respective mailing address for each category where provided.
Telephone_number, email_address: Where provided, contains the observer's contact information.
Affilitation: The observer's affiliation to an organisation, which can be a private company, corporation or non-profit organisation, among others.

Comments:
Send_newsletter: Notes whether the observer wishes to receive newsletters published on NatureWatch.
Password, Password_type, Password_temporary: The observer's password to sign into the website, for access to the data and to enter observations. The temporary password is the one sent automatically to the observer upon registration, where the "password" variable is null unless the observer changed it to a personalised password.
Email_temp: Temporary password automatically sent to the observer upon registration.
Temp_password_sent: This column informs on whether the observer received a confirmation email for his/her registration as an observer for NatureWatch.
Old_observer_code: The database was altered over the years, including a streamlining of the process of assigning numbers to observers. This column contains the numbers for the observers prior to the update, and so more recently added observers do not have a number in this column.
Email_validated: Denotes if the observer received and acknowledged the confirmation email for the creation of his or her account on the NatureWatch website.

1.4 visit

Table: visit		
Column heading	**Data type**	**Reference table**
Pk_visit	Integer – auto-increment	
Fk_observation_location	Integer	Observation_location
Fk_owner	Integer	
Fk_watch_programme	Integer	
Create_date	datetime	
Change_date	datetime	
Visit_date	date	
Visit_date_accuracy_code	Varchar(1)	
Air_temperature	Decimal (6,3)	
Water_temperature	Decimal(6,3)	
comments	Varchar(255)	
C_visit_date	Varchar(30)	
Possible_outlier	Varchar(20)	
validated	Varchar(20)	
Audit_date	date	
Fk_administrator_auditor	Integer	

Pk_visit: Number assigned to the visit of a particular patch of plants (differs from Frogwatch because subsequent visits of the same patch of plants or water body are possible and easy to carry out)
Fk_owner: Defunct column, the foreign key connects nowhere.
Create/change_date: Date where the pk_visit entry was entered or changed (i.e. does not correspond to the actual visit date).
Visit_date: Date where the visit was carried out, as entered by the observer.
Visit_date_accuracy_code: Defunct column, only 338 entries out of over 47 000 are not "NULL".
Air/water_temperature: The air and water temperature entered by the observer, water temperature only where relevant (table is also used in Icewatch).
Comments:

C_visit_date: Serves as a confirmation of the visit date. The purpose of this variable is unclear, as most of the entries have a NULL value.
Possible_outlier: Identifies whether the entry is an outlier. 60 entries out of over 44 000 are considered as such.
Validated: Only 2 entries in this column are labeled "true" in this column, the vast majority being "NULL" with a number of "false" entries.
Audit_date: Defunct column – the only non-"NULL" value being for a "test" entry.
Fk_administrator_auditor: Defunct column.

1.5 Soil_type

Table: soil_type		
Column heading	**Data type**	**Reference table**
Pk_soil_type	Integer – auto-increment	
English_name		Observation_location
French_name		

Pk_soil_type: Code assigned to the soil type.
English/French_name: Name of the soil type, in either language.

1.6 Soil_colour

Table: soil_type		
Column heading	**Data type**	**Reference table**
Pk_soil_colour	Integer – auto-increment	
English_name		Observation_location
French_name		

Pk_soil_colour: Code assigned to a particular soil colour.
English/French_name: Name of the colour, in English or French.

1.7 Worm_species

Table: soil_type		
Column heading	**Data type**	**Reference table**
Pk_worm_species	Integer – auto-increment	
Scientific_name		Observation_location

Pk_worm_species: Code assigned to a worm species.
Scientific_name: Scientific name of the species.

1.8 partner

Table: partner		
Column heading	**Data type**	**Reference table**
Pk_partner	Integer	
Create_date	Datetime	
Change_date	Datetime	
English_name	Varchar(255)	
French_name	Varchar(255)	
Mailing_address_street	Varchar(255)	
Mailing_address_city	Varchar(64)	
Mailing_address_province_code	Varchar(2)	
Mailing_address_country_code	Varchar(3)	
Mailing_address_postal_code	Varchar(6)	
Telephone_number	Varchar(10)	
Fax_number	Varchar(10)	
Email_address	Varchar(128)	
Contact_name	Varchar(128)	
Comments	Varchar(255)	

Pk_partner: Number code assigned to the program partner.
Create/change_date: Date when the pk_partner entry was entered or its details changed.
English/French_name: English and French names of the partner organisation.
Mailing_address_street/city/province_code/country_code/postal_code: Coordinates of the partner.
Telephone/fax_number, email_address: Contact information.
Comments: Personalised comment if any.

It must be noted that there are only 4 entries in this table, none of which contain any information besides the pk_partner and create/change date. The table as a whole is thus defunct.

1.9 age_range

Table: age_range		
Column heading	**Data type**	**Reference table**
Pk_age_range	Integer	
Range_text	Varchar(50)	

Pk_age_range: Assigns an increment code to the age of the observer.
Range_text: The values (high and low) delimiting the above increments.

Appendix E: Other Naturewatch tables

The tables in this section did not connect directly to any of the modules. As a result they are listed alphabetically for convenience, and also because there is no other hierarchy strictly speaking. Most of these tables are administrative or lookup tables for reference by the admins, and thus have little value to the end user.

1.1 Administrator

Table: administrator		
Column heading	**Data type**	**Reference table**
username	Integer	
Fk_watch_programme	Datetime	
superuser	Datetime	
password	Varchar(255)	
first_name	Varchar(255)	
Last_name	Varchar(255)	
province	Varchar(64)	
email	Varchar(2)	
phone	Varchar(3)	
affiliation	Varchar(6)	

Username: The administrator's username entry
Superuser:
Password: The administrator's password.
First/last_name: First and last name of the administrator, where provided (often admin accounts are shared, hence providing one name may be misleading).
Province: Province where the administrator is located.
Email/phone: Administrator's contact information
Affiliation: Any group affiliation of the administrator.

1.2 admin_login_attempts

Table: admin_login_attempts		
Column heading	**Data type**	**Reference table**
User_name_entered	Integer	
Attempted_at	Datetime	
success	Datetime	
Ip_address	Varchar(255)	

User_name_entered: Username entered by the admin for login.
Attempted_at: Date of the attempted login by the admin.
Success: Binary code for whether the login was successful or not.
Ip_address: Admin's ip address for a given login attempt.

1.3 Content

Table: content		
Column heading	**Data type**	**Reference table**
Page_id	Integer	
Description_en	Datetime	
Description_fr	Datetime	
Title_en	Varchar(255)	
Title_fr	Varchar(255)	
Body_en	Varchar(255)	
Body_fr	Varchar(64)	

Page_id: Identifies the page referred to.
Description_en: Description of the page in English.
Description_fr: Description of the page in French
Title_en/fr: English/French title of the page.
Body_en/fr: English/French body text for the page described

1.4 qc_codes

Table: qc_codes		
Column heading	**Data type**	**Reference table**
Table_name	Integer	
Column_name	Datetime	
Qc_codes	Datetime	
name	Varchar(255)	
Create_date	Varchar(255)	
Change_date	Varchar(255)	
description	Varchar(64)	
comments	Varchar(2)	

Table/column_name: Name of the table and column
Qc_codes: One or two-letter code. Is not used anywhere else.
Name: Name of the code entry.
Create/change_date: Date at which the row was entered or modified.
Description: Short description of the row.
Comments: Comments describing the row.

Note that this table only has 4 rows, none of which refer to any other table. The whole table as a whole is thus defunct.

1.5 taxonomy

Table: taxonomy		
Column heading	**Data type**	**Reference table**
Pk_taxon	Integer	
Fk_parameter	Datetime	

genus	Datetime	
species	Varchar(255)	
Create_date	Varchar(255)	
Change_date	Varchar(255)	

Pk_taxon: Corresponds to the frog_id value in the frog_species table.
Genus/Species: Scientific species identifiers
Create/change_date: Date where the table entry was entered/edited.

1.6 validate_email

Table: validate_email		
Column heading	**Data type**	**Reference table**
Pk_validate	Integer	
Fk_observer	Datetime	
email	Datetime	
Hash1	Varchar(255)	
Hash2	Varchar(255)	

Pk_validate: Simple number code unique to the table.
Email: Email address of the observer.
Hash1: Lengthy code – indeterminate use.
Hash2: Lengthy code – indeterminate use.

1.7 valid_parameter_code

Table: valid_parameter_code		
Column heading	**Data type**	**Reference table**
Fk_parameter	Integer	
Parameter_code	Datetime	
Create_date	Datetime	
Change_date	Varchar(255)	
English_name	Varchar(255)	
French_name	Varchar(255)	
English_description	Varchar(64)	
French_description	Varchar(2)	
Fk_watch_programme		

Parameter_code: Code which corresponds to the observation code on a 1to 1 basis, except the parameter_code value is always one point under the frogwatch_observation_code value.
Create/change_date: Date when the entry's information was created/edited.
English/French_name: Name of the entry, in English and French.
English/french_description: A short description of the entry's meaning, in English and French.

This table seems to be largely defunct as it is mostly a repetition of the frogwatch_observation_code table's values over and over, basically for each species (which is largely redundant to do).